

# Thermal Contact I : Symmetries ruled by Exchange Entropy Variations

F. Cornu

Laboratoire de Physique Théorique, UMR 8627 du CNRS  
Université Paris-Sud, Bât. 210  
F-91405 Orsay, France

M. Bauer

Institut de Physique Théorique de Saclay \* CEA Saclay  
F-91191 Gif-sur-Yvette Cedex, France

February 15, 2013

## Abstract

Thermal contact in permanent regime is the archetype of non-equilibrium stationary states driven by constant non-equilibrium constraints enforced by reservoirs of exchanged conserved microscopic quantities. In the present paper we are interested in a class of models with a finite number of possible configurations where the heat exchanges are described as changes in the populations of energy levels when the configurations stochastically evolve under a master equation. Systems with discrete configurations have no microscopic Hamiltonian dynamics but, if the microscopic deterministic dynamics which conserves energy is assumed to be ergodic, it is shown that the transition rates obey a modified detailed balance (MDB) which is generically expressed in terms of the microscopic exchange entropy variation, namely the opposite of the variation of the thermodynamic entropy of the reservoirs under a microscopic jump of the system configuration. We investigate the generic statistical properties for measurable quantities which arise from the MDB constraint. For a finite-time evolution (transient regime) from an initial equilibrium state we derive a detailed fluctuation relation for the excess exchange entropy variation and an associated integral fluctuation relation. In the non-equilibrium stationary state (long-time regime), the proper mathematical definition of a large deviation function is introduced together with alternative definitions, and fluctuation relations are derived. The fluctuation relation for the exchange entropy variation is merely a particular case of the Lebowitz-Spohn fluctuation relation for the action functional [1]. The infinite time limit of any odd cumulant per unit time of an exchanged quantity is expressed in terms of a series involving higher even cumulants and powers of the thermodynamic force associated to the mean current ; every relation can be seen as a generalized Green-Kubo relation valid far from equilibrium. Its generalization to the case of several independent macroscopic currents is derived. It entails a relation between the  $n$ th-order non-linear response coefficient of any odd cumulant per unit time in the vicinity of equilibrium and a finite number of lower-order non-linear response coefficients of even cumulants per unit time. The latter relations can be seen as another kind of generalizations of the standard Green-Kubo formula pertaining to the first-order coefficient.

**PACS** : 05.70.Ln, 02.50.Ga, 05.60.Cd

---

\*CEA/DSM/IPhT, Unité de recherche associée au CNRS

**KEYWORDS** : thermal contact; master equation; ergodicity; modified detailed balance ; exchange entropy; large deviation function ; fluctuation relations; generalized Green-Kubo relations.

*Corresponding author* :  
 CORNU Françoise,  
 Fax: 33 1 69 15 82 87  
 E-mail: Francoise.Cornu@u-psud.fr

## Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
1.1	Issues at stake . . . . .	3
1.2	Main results valid beyond the thermal contact example . . . . .	5
1.2.1	Constraints from ergodicity and energy conservation . . . . .	5
1.2.2	Various entropy variations inherent to Markovian dynamics . . . . .	6
1.2.3	MDB and non-perturbative symmetries in transient regimes . . . . .	7
1.2.4	MDB and fluctuation relations in the stationary regime . . . . .	8
1.2.5	MDB and generalized Green-Kubo relations . . . . .	8
<b>2</b>	<b>Constraints upon transitions rates</b>	<b>9</b>
2.1	Ergodicity . . . . .	10
2.1.1	Ergodicity in classical Hamiltonian dynamics . . . . .	10
2.1.2	Ergodicity in deterministic dynamics for discrete variables . . . . .	10
2.1.3	Constraints from ergodicity and interaction pattern upon coarse-grained evolution . . . . .	11
2.2	Markovian approximation for the mesoscopic dynamics . . . . .	12
2.2.1	Definition of a Markovian approximation for the mesoscopic dynamics . . . . .	12
2.2.2	Microcanonical detailed balance and other properties . . . . .	12
2.2.3	Further consequence of the interaction pattern . . . . .	13
2.3	Transient regime when large parts are described in the thermodynamic limit . . . . .	14
2.3.1	Large parts in the thermodynamic limit . . . . .	14
2.3.2	Transient regime and modified detailed balance (MDB) . . . . .	14
2.3.3	Expression of MDB in terms of exchange entropy variation . . . . .	16
<b>3</b>	<b>Master equation, exchange entropy flow and various entropy variations</b>	<b>17</b>
3.1	Evolution of the probability distribution without MDB (known results) . . . . .	17
3.1.1	Generic properties . . . . .	17
3.1.2	Properties arising from the irreducibility condition . . . . .	18
3.2	Microscopic currents . . . . .	19
3.2.1	Definitions . . . . .	19
3.2.2	Probabilistic interpretation of microscopic observable currents . . . . .	19
3.2.3	Mean energy time derivative . . . . .	20
3.2.4	Mean exchange entropy flow . . . . .	20
3.3	Evolution of the Shannon-Gibbs entropy . . . . .	21
3.3.1	Definition of the entropy production rate . . . . .	21
3.3.2	Comparison with the thermodynamics of irreversible processes . . . . .	21
3.3.3	Entropy production rate under MDB . . . . .	22
3.3.4	Decomposition of the entropy production rate into two positive contributions . . . . .	23
3.3.5	Comparison with NESS characterization from graph theory . . . . .	24
<b>4</b>	<b>Exchange entropy variation and symmetries at finite time under MDB</b>	<b>26</b>
4.1	Exchange entropy variation for a history . . . . .	26
4.2	MDB and symmetry between time-reversed histories . . . . .	27
4.3	Symmetry between time-reversed evolutions with fixed heat amounts . . . . .	27
4.4	Symmetries in protocols starting from an equilibrium state . . . . .	28
4.5	Symmetries in protocols starting from a stationary state with a canonical distribution . . . . .	29

<b>5</b>	<b>Long-time symmetries : non-equilibrium steady state with MDB</b>	<b>29</b>
5.1	Fluctuation relations . . . . .	29
5.1.1	Fluctuation relation for the cumulated exchange entropy variation . . . . .	29
5.1.2	Constraints from the bound upon the system energy . . . . .	30
5.1.3	The Gärtner-Ellis theorem and some of its consequences . . . . .	31
5.1.4	Linear response far from equilibrium . . . . .	32
5.1.5	Green-Kubo relation as a consequence of the fluctuation relation for the large deviation function . . . . .	33
5.2	Symmetry of the generating function for cumulants per unit time under MDB . . . . .	34
5.2.1	Infinite-time cumulants per unit time for $\mathcal{Q}_1$ and $\mathcal{Q}_2$ . . . . .	34
5.2.2	Constraints from the bound upon the system energy . . . . .	35
5.2.3	MDB and symmetry of the generating function for infinite-time cumulants per unit time . . . . .	35
5.3	Far from equilibrium relations for infinite-time heat cumulants per unit time . . . . .	37
5.3.1	Generalized Green-Kubo relations for cumulants per unit time far from equilibrium . . . . .	37
5.3.2	Relations between non-linear responses of cumulants per unit time near equilibrium . . . . .	39
<b>6</b>	<b>Extension of previous results to stationary states with several independent currents</b>	<b>41</b>
6.1	Consequences of MDB at finite time . . . . .	41
6.2	Consequences of MDB in the infinite-time limit . . . . .	42
6.3	Generalized Green-Kubo relations for several independent currents . . . . .	43
<b>A</b>	<b>Property of the coarse grained dynamics</b>	<b>44</b>
<b>B</b>	<b>The Markovian approximation</b>	<b>45</b>
B.1	Discrete time stochastic approximation . . . . .	45
B.2	Continuous time approximation . . . . .	46
<b>C</b>	<b>Microcanonical detailed balance and time reversal invariance in Hamiltonian dynamics</b>	<b>47</b>
<b>D</b>	<b>Definitions for statistics over histories</b>	<b>48</b>
<b>E</b>	<b>Some remarks on large deviations</b>	<b>49</b>
E.1	The mathematical definition of large deviations . . . . .	49
E.2	Alternative definitions . . . . .	51
E.3	Inequalities at finite time and fluctuation relation between large deviation functions . . . . .	52
E.4	Cumulated quantities with sub-extensive difference . . . . .	53

# 1 Introduction

## 1.1 Issues at stake

Thermal contact between two energy reservoirs is one of the first issues addressed by early thermodynamics and its second principle about entropy variation. It has been revisited in the context of the description of transport phenomena either from the statistical approach by the Boltzmann equation and its BBGKY hierarchy generalization or from the phenomenological thermodynamics of irreversible processes [2, 3, 4] and the local equilibrium phenomenological approach for inhomogeneous continuous media [5]. Until recently there have been repeated attempts to formulate thermodynamics for non-equilibrium steady states [6, 7]. Nowadays the statistical description for the fluctuations of heat exchanges between a small system and two thermal baths is part of the so-called “stochastic thermodynamics” of small systems incorporating the effects of fluctuations [8, 9]. In the last two decades the latter domain has been the subject of an increasingly intense activity, from the theoretical as well as from the experimental point of view with very fast technological improvements [10, 11]. (For a pedagogical introduction in a nutshell see for instance Ref.[12]. A recent extended review is to be found in Ref.[9]).

First steps in out-of-equilibrium statistical mechanics have been the study of the linear (static or dynamic) response to an external constraint that drives the system out of its equilibrium state. This vast topic began with the Einstein fluctuation-dissipation relation for Brownian motion (see collected translated papers in [13]) and is still active (for a review see [14]). Nowadays, the question of linear response in the vicinity of a non-equilibrium steady states is under investigation (see for instance [15, 16, 17, 18]), but the subject is beyond the scope of the present work.

Non-perturbative approaches for systems far from equilibrium have followed various ways. One trend has been the search for a unifying variational principle based on a large deviation point of view, which would generalize the entropy-probability relation introduced by Einstein in his theory of thermodynamic equilibrium fluctuations [19] in order to explain the opalescence phenomenon. Indeed, principles of equilibrium statistical mechanics can be seen as resulting from the maximization of the Shannon entropy under various constraints about macroscopic observables in the framework of information theory [20, 21]. From this point of view, statistical mechanics, which allows one to retrieve the laws of thermodynamics while describing mesoscopic fluctuations, may be interpreted as an example of a large deviation theory, which provides the probabilities for an observable to deviate from its most probable value [22]. In order to go beyond the phenomenological irreversible thermodynamics, Oono [23] has promoted the idea of applying the large deviation theory to the non-equilibrium statistical physics seen as a statistical mechanics along the time axis. In this direction efforts have been devoted to the study of the maximization of the Shannon entropy for the system histories under some constraints, such as fixed values for the macroscopic out-of-equilibrium currents in a non-equilibrium steady state (see [24, 25] and references therein).

In the same spirit a second path, the quest for some relevant physical quantities that would obey some universal principle, is the subject of the fluctuation relations either at finite time (transient regimes) or in the infinite-time limit (stationary regimes). The active topic of fluctuation relations was initiated by works about a symmetry of the large deviation function of some entropy production rate in the non-equilibrium stationary states of some chaotic dynamical systems [26, 27, 28, 29], and about the Jarzynski identity for the finite-time cumulated work when a Hamiltonian system is driven out of an initial equilibrium state by the variation of some external parameter [30]. The first results pertained to classical systems but fluctuation relations in quantum systems have also been investigated (see for instance the reviews [31, 32]). A short list for successive steps on the narrower pathway of fluctuation relations for systems with Markovian stochastic dynamics is given by the following references [33, 34, 35, 36, 1, 37, 38, 8, 39]. Among these results, in the more specific case of systems with stochastic evolutions described by a master equation for the probability of the system configurations, the milestone is the fluctuation relation obeyed by the dimensionless action functional introduced by Lebowitz and Spohn [1].

In the even more restricted case where the system exchanges conserved microscopic quantities, such as energy quanta or particles or elementary volumes, with infinite size reservoirs which stay in their equilibrium states during the considered evolution of the finite system, the transition rates must obey a relation, which will be referred to as the Modified Detailed Balance (MDB)<sup>1</sup>. When the transition rates obey the MDB, the action functional introduced by Lebowitz and Spohn coincides with the variation of the sum of the reservoirs thermodynamic entropies. Then, apart from the fluctuation relation obeyed by the latter entropy variation, the known results about large deviation functions which are the most relevant for the topic of the present paper deal with the following quantities in finite size systems : the heat current which goes through the finite system which sets up a thermal contact between two thermostats [40, 41] (also first addressed in the context of classical or quantum Hamiltonian dynamics in Ref.[42]); more generally the various macroscopic currents through a system in a non-equilibrium stationary state sustained by various kinds of reservoirs [43, 44].

Apart from general approaches, a third trend in the quest for some non-perturbative statistical theory of out-of-equilibrium phenomena has been the search for solvable models which could give some hints in the comprehension of these phenomena in the absence of any theoretical framework. We are interested in a class of models where the system has a finite number of possible configura-

---

<sup>1</sup>The denomination “generalized” detailed balance is also to be found in the literature.

tions, the heat exchanges are described as changes in the populations of its energy levels, and the configurations stochastically evolve under a master equation with transition rates bound to obey the MDB relation arising from the existence of an underlying ergodic deterministic microscopic dynamics which conserves energy. In paper II [45] we perform explicit analytical calculations for the very simple case where the thermal contact system is reduced to two spins, each of which is flipped by a single thermostat, the two thermostats being at different temperatures.

## 1.2 Main results valid beyond the thermal contact example

In the present section we point out the main ideas and generic results of our study when the system is driven out of equilibrium by exchanges of some conserved microscopic quantities between the system and reservoirs characterized by constant thermodynamic intensive parameters. For the sake of pedagogy, we express them in the language of the thermal contact example. The general arguments and results in the situation where there are more than two reservoirs and/or several kinds of exchanged quantities are more detailed in section 6.

In the whole paper we put emphasis on the exchange entropy variation  $\Delta_{\text{exch}}S$  of the system, which is opposite to the well-defined entropy variation of the reservoirs described in the thermodynamic limit, because it seems to be the crucial quantity to consider in order to answer the key-point question of identifying the relevant measurable quantities which are to obey non-trivial universal properties (beyond long-time decorrelations and the subsequent existence of large deviation functions). This point of view slightly differs from the common picture where the focus is put on the variation of the Shannon-Gibbs entropy of the system. The latter picture has arisen from the seminal works by Crooks [35, 36, 46] and has led to the notion of entropy production along a stochastic trajectory of the system microscopic configurations (see [38, 9] for a recent formulation). In the whole paper Boltzmann constant is set to 1 and  $S$  denotes dimensionless entropies.

We point out that Appendix E contains the proper mathematical definition of the large deviation function, and some variants for the models at stake, together with demonstrations of properties about large deviation functions which are not usually exhibited in the literature of the physicists community.

### 1.2.1 Constraints from ergodicity and energy conservation

We consider models for thermal contact where the energy exchanges between several very large systems with indices  $a$ 's are realized through interactions with a system which has a finite number of configurations, namely which has discrete variables in finite number. The latter system is referred to as the “spin system” in the following. The spin system has an internal energy which is bounded, since its configuration space is finite. When each large part with index  $a$  is described only at a macroscopic level, i.e. only by its energy  $E_a$ , we have to resort to a statistical description, where the knowledge of the system is specified by the following datas : the energies of the large parts and the configuration  $\mathcal{C}$  of the spin system. Such a set of datas defines what we call a mesoscopic configuration of the total system in the sequel. Then the stochastic evolution is determined by transition rates between these mesoscopic configurations. The key points are the following.

First we have to answer the question : which constraints must be obeyed by the choice for the transition rates ? Discrete variables cannot evolve under a microscopic Hamiltonian dynamics, but the existence of equilibrium in the infinite-time limit is ensured if the deterministic microscopic dynamics is ergodic and conserves the sum of the energies of every part. In a given energy level all ergodic microscopic dynamics have the same period and the interaction pattern entails that, during a period of the microscopic dynamics, the number of transitions which occur from one mesoscopic configuration of the total system to another one is equal to the number of transitions in the opposite sense (subsection 2.1 and Appendix A). As a consequence, if the mesoscopic dynamics is approximated by a Markov process according to the prescription derived in Appendix B, then the corresponding transition rates must satisfy three constraints (subsection 2.2). In particular the transition rates at the mesoscopic level obey the microcanonical detailed balance

(2.9). (For a comparison, the derivation of the microcanonical detailed balance in the case of continuous mesoscopic variables and an underlying microscopic Hamiltonian dynamics invariant under time reversal is presented in Appendix C). The interaction pattern further transforms the microcanonical balance into the reduced expression (2.11) which involves the Boltzmann entropy variation of the large part which exchanges energy in the transition.

In the limit where the sizes of the large parts go to infinity before the time evolution of the system is considered, the global system does not reach equilibrium but the large parts are in the thermodynamic equilibrium state which they would tend to in the infinite-time limit if they were isolated (subsection 2.3). In the case of thermal contact the large parts behave as thermostats characterized by their inverse temperatures  $\beta_a$ 's. Then the previous constraints for the transition rates between mesoscopic configurations become three conditions which must be obeyed by the transition rates between the spin system configurations in the corresponding transient regime. Let  $(\mathcal{C}'|\mathbb{W}|\mathcal{C})$  denote the transition rate from configuration  $\mathcal{C}$  to configuration  $\mathcal{C}'$  in the transient regime. The three conditions are the following ones. First, the set of transition rates must be such that any configuration  $\mathcal{C}$  can be reached by a succession of jumps with non-zero transition rates from any configuration  $\mathcal{C}'$ , namely in the network representation of the stochastic evolution

$$\text{the graph } G \text{ associated with the transition rates is connected.} \quad (1.1)$$

In other words the Markov matrix  $\mathbb{M}$  defined from the transition rates by (2.15) must be irreducible and the property (1.1) will be referred to as the irreducibility condition. Second, the transition rates must obey the microscopic reversibility condition for any couple of configurations  $(\mathcal{C}, \mathcal{C}')$ , namely

$$(\mathcal{C}'|\mathbb{W}|\mathcal{C}) \neq 0 \quad \Leftrightarrow \quad (\mathcal{C}|\mathbb{W}|\mathcal{C}') \neq 0. \quad (1.2)$$

Third, the transition rates have to satisfy the modified detailed balance (MDB) which takes the form (2.16) in the case of thermal contact.

We choose to reexpress the MDB relation between transition rates in terms of the exchange entropy variation  $\delta_{\text{exch}} S(\mathcal{C}' \leftarrow \mathcal{C})$  associated with a jump of the spin system from a microscopic configuration  $\mathcal{C}$  to another one  $\mathcal{C}'$ :  $\delta_{\text{exch}} S(\mathcal{C}' \leftarrow \mathcal{C})$  is the opposite of the microscopic variation of the sum of the thermodynamic entropies of the heat baths when a single thermostatchange the spin system energy by flipping a spin. The MDB relation (2.16) for the thermal contact reads quite generally

$$\frac{(\mathcal{C}'|\mathbb{W}|\mathcal{C})}{(\mathcal{C}|\mathbb{W}|\mathcal{C}')} = e^{-\delta_{\text{exch}} S(\mathcal{C}' \leftarrow \mathcal{C})}. \quad (1.3)$$

This formulation of the MDB is the most generic statement, because it also applies to exchanges of other kinds of conserved microscopic quantities, such as particle numbers or elementary volumes, and it may involve simultaneous exchanges of different kinds of conserved quantities. We stress that  $\delta_{\text{exch}} S(\mathcal{C}' \leftarrow \mathcal{C})$  is defined unambiguously and involves experimentally measurable quantities (in the thermal contact case the inverse temperatures of the energy reservoirs  $\beta_a$ 's as well as the heat amounts received by the system).

### 1.2.2 Various entropy variations inherent to Markovian dynamics

In section 3 we stress some consequences of the fact that the configuration probability obeys a master equation. First, even if the transition rates do not obey the MDB, the irreducibility of the Markov matrix implies the uniqueness of the stationary state and the role played by the relative entropy with respect to the stationary solution,  $S_{\text{rel}}[P(t)|P_{\text{st}}]$  is recalled (subsection 3.1). We introduce the generic currents (3.10)-(3.11) for a microscopic variation when the system goes out of a given configuration  $\mathcal{C}$ ; the averages of such currents show off either in the mean energy time derivative or in the mean exchange entropy flow (subsection 3.2). In order to make the comparison with the thermodynamics of irreversible processes, the time-derivative of the Shannon-Gibbs entropy  $S^{SG}[P(t)]$  of the system is split into two contributions, an exchange (or external) part  $d_{\text{exch}} S/dt$  arising from exchanges with the external reservoirs, hereafter called the exchange entropy flow, and an internal (or irreversible) part  $d_{\text{int}} S^{SG}/dt$  due to the internal irreversible

processes in the system, also called the entropy production rate (subsection 3.3). In the stationary state, the time derivative  $dS^{SG}[P_{\text{st}}]/dt$  vanishes and the entropy production rate becomes opposite to the exchange entropy flow. Henceforth, in the case where there are only two reservoirs of energy, the stationary entropy production rate can be rewritten as in phenomenological irreversible thermodynamics, namely

$$-\left.\frac{d_{\text{exch}}S}{dt}\right|_{\text{st}} = \left.\frac{d_{\text{int}}S^{SG}}{dt}\right|_{\text{st}} = \mathcal{F}J \quad (1.4)$$

where  $J$  is the mean heat current which goes through the system from the heat bath 2 to the heat bath 1 and  $\mathcal{F} = \beta_1 - \beta_2$  is the associated “thermodynamic force”. When the MDB is satisfied the expressions for  $d_{\text{exch}}S/dt$  and  $d_{\text{int}}S^{SG}/dt$  coincide with the splitting of  $dS^{SG}/dt$  introduced in Ref.[1]. Moreover, the entropy production rate appears as the sum of two positive contributions : on the one hand, the opposite of the time-derivative of the relative entropy  $S_{\text{rel}}[P(t)|P_{\text{st}}]$  and, on the other hand, the average of a microscopic current associated to some stationary affinity variation when the system jumps out of a configuration. The latter current proves to be positive even before averaging over the configurations (see (3.40)). For the sake of completeness we recall how, in the graph theory where the master equation is represented by a network, a non equilibrium stationary state is characterized by the affinities and probability currents associated with a restricted number of cycles defined from the network. When the graph is a pure cycle, there is a single affinity which has a simple probabilistic interpretation given in paper II [45].

### 1.2.3 MDB and non-perturbative symmetries in transient regimes

In the form (1.3) where it involves the microscopic exchange entropy variation  $\delta_{\text{exch}}S(\mathcal{C}' \leftarrow \mathcal{C})$  associated with a jump from configuration  $\mathcal{C}$  to configuration  $\mathcal{C}'$ , the MDB entails time-reversal symmetries for finite time intervals at more and more mesoscopic levels as follows. The ratio (4.6) of the probabilities for a microscopic history  $\mathcal{H}ist$  and the time-reversed one is determined by the cumulated exchange entropy variation  $\Delta_{\text{exch}}S[\mathcal{H}ist]$  along a history defined in (4.1). Then, the ratio (4.8) between, on the one hand, the probability for all evolutions with fixed initial and final configurations and given heat amounts  $\mathcal{Q}_1$  and  $\mathcal{Q}_2$  received from the thermostats and, on the other hand, the probability for all backward evolutions with exchanged initial and final configurations and opposite heat amounts  $-\mathcal{Q}_1$  and  $-\mathcal{Q}_2$  is determined by the exchange entropy variation

$$\Delta_{\text{exch}}S(\mathcal{Q}_1, \mathcal{Q}_2) = \beta_1 \mathcal{Q}_1 + \beta_2 \mathcal{Q}_2. \quad (1.5)$$

When the system is prepared in an equilibrium state at the inverse temperature  $\beta_0$  and suddenly put in thermal contact at the initial time of measurements with two heat baths at the inverse temperatures  $\beta_1$  and  $\beta_2$ , the symmetry (4.8) together with the specific form of the equilibrium canonical distribution lead one to consider the following measurable quantity : the excess exchange entropy variation  $\Delta_{\text{exch}}^{\text{excs},\beta_0}(\mathcal{Q}_1, \mathcal{Q}_2)$  defined as the difference between the exchange entropy variation under the non-equilibrium external constraint  $\beta_1 \neq \beta_2$  and that under the equilibrium condition  $\beta_1 = \beta_2 \equiv \beta_0$ , namely

$$\Delta_{\text{exch}}^{\text{excs},\beta_0}(\mathcal{Q}_1, \mathcal{Q}_2) = \Delta_{\text{exch}}S(\mathcal{Q}_1, \mathcal{Q}_2) - \beta_0(\mathcal{Q}_1 + \mathcal{Q}_2). \quad (1.6)$$

$\Delta_{\text{exch}}^{\text{excs},\beta_0}(\mathcal{Q}_1, \mathcal{Q}_2)$  obeys the symmetry relation at any finite time, or “detailed fluctuation relation”,

$$\frac{P_{P_{\text{can}}^{\beta_0}}\left(\Delta_{\text{exch}}^{\text{excs},\beta_0}S\right)}{P_{P_{\text{can}}^{\beta_0}}\left(-\Delta_{\text{exch}}^{\text{excs},\beta_0}S\right)} = e^{-\Delta_{\text{exch}}^{\text{excs},\beta_0}S}. \quad (1.7)$$

The latter relation itself entails the identity, or “integral fluctuation relation”,  $\langle e^{\Delta_{\text{exch}}^{\text{excs},\beta_0}S} \rangle_{P_{\text{can}}^{\beta_0}} = 1$  in the spirit of Jarzynski identity [30]. To our knowledge these two relations have not appeared explicitly in the literature, though they may be seen as a transposed application of the argument first exhibited by Crooks [36] and then Seifert [38, 8] for the entropy production along a stochastic

trajectory when the system is in thermal contact with only one heat bath and is driven out of equilibrium by a time-dependent external parameter. We notice that if the configuration probability distribution in a stationary state of the system happens to be the canonical distribution at an effective inverse temperature  $\beta_*(\beta_1, \beta_2)$ , as it is the case in the solvable model considered in paper II [45], there exist similar detailed and integral fluctuation relations for a protocol where the system is initially prepared in a non-equilibrium stationary state with two heat baths at the inverse temperatures  $\beta_1^0$  and  $\beta_2^0$  and suddenly put at the initial time of measurements in thermal contact with two heat baths at the inverse temperatures  $\beta_1$  and  $\beta_2$ . Then  $\Delta_{\text{exch}}^{\text{excs}, \beta_0}(\mathcal{Q}_1, \mathcal{Q}_2)$  is to be replaced by  $\Delta_{\text{exch}}^{\text{excs}, \beta_*^0}(\mathcal{Q}_1, \mathcal{Q}_2)$  where  $\beta_*^0$  is the effective inverse temperature  $\beta_*^0(\beta_1^0, \beta_2^0)$ .

#### 1.2.4 MDB and fluctuation relations in the stationary regime

For a system with a finite number of configurations, when the Markov stochastic matrix for the continuous-time evolution of the configuration probabilities is irreducible, (see definition (1.1)), the Perron-Frobenius theorem can be applied: the system has a single stationary state, and it is such that every configuration has a non-vanishing weight (see for instance Ref. [47]). Moreover the system reaches its stationary state in an exponentially-short time [48]. Then the symmetry relation (4.8) enforced by the MDB at finite time leads to the existence of lower and upper bounds for the ratio between the finite-time probability to measure heat amounts  $\mathcal{Q}_1$  and  $\mathcal{Q}_2$  and the corresponding probability for the opposite values, when the system is in its stationary state (see (5.1)). Similar bounds are exhibited in Ref.[41].

The Markovian property of the evolution implies that the cumulated heats  $\mathcal{Q}_1$  and  $\mathcal{Q}_2$ , and subsequently the cumulated exchange entropy variation  $\Delta_{\text{exch}}S$ , all grow linearly with time in the long-time limit and that there exist large deviation functions for each of them [49, 50]. The proper mathematical definition of a large deviation function is recalled in Appendix E.1, and other alternative definitions when exchanged quantities are discrete are introduced in Appendix E.2.

The finite-time inequalities (5.1) entail a symmetry in the long-time limit, according to the general results derived in Appendix E.3 : the dimensionless exchange entropy variation  $\Delta_{\text{exch}}S$  obeys the fluctuation relation

$$f_{\Delta_{\text{exch}}S}(\mathcal{J}) - f_{\Delta_{\text{exch}}S}(-\mathcal{J}) = -\mathcal{J}, \quad (1.8)$$

where  $\mathcal{J}$  denotes the values taken by the cumulated current  $\Delta_{\text{exch}}S(t)/t$ . The fluctuation relation for  $\Delta_{\text{exch}}S$  is a special case of the more general fluctuation relation for the action functional introduced by Lebowitz and Spohn [1] ; indeed when the transition rate obey the MDB, the action functional for a given history becomes equal to the entropy variation of the heat reservoirs, namely to the opposite of the exchange entropy variation.

For a system with a finite number of configurations,  $\mathcal{Q}_1 + \mathcal{Q}_2$  is bounded. Then, according to the results of Appendix E.4, the large deviation function for  $\mathcal{Q}_2$  is equal to that for  $-\mathcal{Q}_1$ , while the fluctuation relation (1.8) for  $f_{\Delta_{\text{exch}}S}$  also entails a fluctuation relation for the large deviation function  $f_{\mathcal{Q}_2}$  for the cumulated heat  $\mathcal{Q}_2$ , because the difference between  $\Delta_{\text{exch}}S$  and  $-(\beta_1 - \beta_2)\mathcal{Q}_2$  is finite at any time  $t$ . The fluctuation relation can be written in the generic form

$$f_{\mathcal{Q}_2}(\mathcal{J}) - f_{\mathcal{Q}_2}(-\mathcal{J}) = \mathcal{F}\mathcal{J}, \quad (1.9)$$

where  $\mathcal{F}$  is the thermodynamic force conjugated to the mean current  $J = \lim_{t \rightarrow +\infty} \langle \mathcal{J}(t) \rangle = \lim_{t \rightarrow +\infty} \langle \mathcal{Q}_2(t) \rangle / t$  through the expression of the exchange entropy flow in the stationary state (see (1.4)).

#### 1.2.5 MDB and generalized Green-Kubo relations

Quite generally, when there exists a large deviation function for the cumulated current  $\mathcal{J}_t \equiv X_t/t$  associated with the cumulated quantity  $X_t$ , the generic expression of the linear response in the non-equilibrium state far from equilibrium is given in (5.20) : it relates on the one hand, the coefficient  $\partial J / \partial \mathcal{F}$  of the linear response of the heat current  $J$  to a variation of the thermodynamic



force  $\mathcal{F}$  and, on the other hand, the infinite-time limit of the variance per unit time for the cumulated quantity  $X_t$  in the non-equilibrium steady state, with a coefficient whose expression depends on the system.

In the limit where the thermodynamic force  $\mathcal{F}$  vanishes,  $\partial J/\partial \mathcal{F}$  tends to the linear-response coefficient near equilibrium, namely the kinetic Onsager coefficient,  $L \equiv \partial J/\partial \mathcal{F}|_{\mathcal{F}=0}$ . If the system obeys the fluctuation relation (1.9), with  $X_t$  in place of  $\mathcal{Q}_2$ , then the coefficient in the identity (5.20) takes the universal value  $\frac{1}{2}$  and the identity becomes the fluctuation-dissipation relation (also referred to as the Green-Kubo relation) between the kinetic Onsager coefficient and the infinite-time limit of the variance per unit time of  $X_t$  at equilibrium. In the case of thermal contact, in the limit where the difference  $\beta_1 - \beta_2$  between the inverse temperatures vanishes, the ratio between the stationary heat current which goes through the system from the heat bath 2 to the heat bath 1 and the difference  $\beta_1 - \beta_2$  becomes equal to  $\frac{1}{2}$  times the variance per unit time of the heat amount exchanged with one thermal bath at equilibrium. (When the total system is at equilibrium, the net heat amount  $\mathcal{Q}_1 + \mathcal{Q}_2$  received by the system remains finite at any time, but the heat amount received from one bath grows linearly in time in the long-time limit).

For a system with a finite number of configurations, the MDB entails a symmetry upon the generating function for the infinite-time limit of the cumulants of  $X_t$  per unit time. This symmetry takes the generic form (5.54) in terms of the thermodynamic force  $\mathcal{F}$ . We then show that, in the infinite-time limit, any odd cumulant per unit time  $\kappa^{[2n+1]}/t$  can be expressed in terms even-order cumulants per unit time through the relation

$$\lim_{t \rightarrow +\infty} \frac{\kappa^{[2n+1]}(\mathcal{F})}{t} = \sum_{k=0}^{+\infty} d_k \mathcal{F}^{2k+1} \lim_{t \rightarrow +\infty} \frac{\kappa^{[2(n+k+1)]}(\mathcal{F})}{t} \quad \text{for } n = 0, 1, \dots, \quad (1.10)$$

where  $d_k$  is given in (5.64). The corresponding expressions for the ratios  $(1/\mathcal{F}) \times \lim_{t \rightarrow +\infty} \kappa^{[2n+1]}/t$  may be viewed as generalized Green-Kubo relations valid both far from equilibrium and for all cumulants.

We also express the response of any odd cumulant per unit time at any order in the thermodynamic force  $\mathcal{F}$  near equilibrium in terms of non-linear response coefficients for even cumulants per unit time at lower orders in  $\mathcal{F}$  near equilibrium (see (5.70)). At first order one gets the generalized fluctuation-dissipation relations (5.71) for any odd cumulant.

In the more general situation where there are several independent mean currents between reservoirs, we derive the corresponding generalized Green-Kubo relations (6.21). From the latter equations one can derive relations between non-linear response coefficients. These relations have already been settled by another method by Andrieux and Gaspard [51]. As noticed by these authors some of them lead to symmetries which are generalizations of the Onsager reciprocity relations.

## 2 Constraints upon transitions rates

In the present section we review some of the constraints that ergodic deterministic energy-conserving microscopic dynamics puts on the statistical mesoscopic description of a finite system which establishes thermal contact between energy reservoirs.

Indeed, the following situation occurs commonly : the interactions in the whole system allow to define one small part in contact with otherwise independent large parts. The large parts, which involve a huge number of degrees of freedom, do not interact directly among each other (this gives a criterion to identify the distinct large parts), but are in contact with the small part, which involves only a few degrees of freedom. Moreover each degree of freedom in the small part is directly in contact with at most one large part. This results in a star-shaped interaction pattern. It is convenient then to forget about the microscopic description of the large parts, and turn to a statistical description of their interactions with the small part. Some general features of the statistical description can be inferred from microscopic ergodicity.

## 2.1 Ergodicity

### 2.1.1 Ergodicity in classical Hamiltonian dynamics

In classical mechanics, the time evolution of a system in phase space is described by a Hamiltonian  $H$ . If the system is made of several interacting parts, the Hamiltonian splits as  $H = H_{\text{dec}} + H_{\text{int}}$ , where  $H_{\text{dec}}$  accounts for the dynamics if the different parts were decoupled and  $H_{\text{int}}$  accounts for interactions. The energy hypersurface  $H = E$ , usually a compact set, is invariant under the time evolution, and in a generic situation, this will be the only conserved quantity.

The ergodic hypothesis states that a generic trajectory of the system will asymptotically cover the energy hypersurface uniformly. To be more precise, phase space is endowed with the Liouville measure (i.e. in most standard cases the Lebesgue measure for the product of couples made by every coordinate and its conjugate momentum), which induces a natural measure on the energy hypersurface, and ergodicity means that in the long run, the time spent by the system in each open set of the energy hypersurface will be proportional to its Liouville measure. Ergodicity can sometimes be built in the dynamics, or proved, but this usually requires immense efforts.

Ergodicity depends crucially on the fact that the different parts are coupled : if  $H_{\text{int}} = 0$ , each part will have its energy conserved, and motion will take place on a lower-dimensional surface. If  $H_{\text{int}}$  is very small, the system will spend a long time very nearby this lower-dimensional surface, but at even longer time scales, ergodicity can be restored. By taking limits in a suitable order (first infinite time and then vanishing coupling among the parts) one can argue that the consequences of ergodicity can be exploited by reasoning only on  $H_{\text{dec}}$ .

Notice that in this case, we have in fact some kind of dichotomy :  $H_{\text{dec}}$  defines the energy hypersurface, but cannot be used to define the ergodic motion, which is obtained from  $H_{\text{dec}} + H_{\text{int}}$  via a limiting procedure. So the dynamics conserves  $H_{\text{dec}}$  but is not determined by  $H_{\text{dec}}$ .

### 2.1.2 Ergodicity in deterministic dynamics for discrete variables

Our aim is to translate the above considerations in the context of a large but finite system described by discrete variables such as classical Ising spins.

In the case of discrete dynamical variables, one can still talk about the energy  $E$  of a configuration, but there is no phase space and no Hamiltonian dynamics available. So there is no obvious canonical time evolution. This is where we exploit the previously mentioned dichotomy: we do not define the time evolution in terms of  $E$ , but simply impose that the deterministic time evolution preserves  $E_{\text{dec}}$  and that it respects the star-shaped interaction pattern between the small part and the large parts.

Ergodicity is most simply expressed if one assumes that time is discrete as well. Then deterministic dynamics is given by a bijective map, denoted by  $\mathcal{T}$  in what follows, on configuration space, applied at each time step to get a new configuration from the previous one. As the configuration space is finite, the trajectories are bound to be closed. Then, for a given initial value  $E$  of  $E_{\text{dec}}$ , a specific dynamics  $\mathcal{T}$  conserving  $E_{\text{dec}}$  corresponds to a periodic evolution of the microscopic configuration of the full system inside the energy level  $E_{\text{dec}} = E$ .

Ergodicity entails that the corresponding closed trajectory covers fully the energy level  $E_{\text{dec}} = E$ , and then it must cover it exactly once during a period because the dynamics is one to one. As a consequence the period of the ergodic evolution inside a given energy level of  $E_{\text{dec}}$  is the same for all choices of ergodic dynamics  $\mathcal{T}$  which conserves  $E_{\text{dec}}$ . This period, denoted by  $N$  in time step units, depends only on the value  $E$  of  $E_{\text{dec}}$ ,

$$N = \Omega_{\text{dec}}(E), \quad (2.1)$$

where  $\Omega_{\text{dec}}(E)$  denotes the total number of microscopic configurations in the level  $E_{\text{dec}} = E$ . This is reminiscent of the microcanonical ensemble. Let us note that in classical mechanics, there is a time reversal symmetry, related to an involution of phase space, changing the momenta to their opposites while leaving the positions fixed (see (C.2)). In the discrete setting, involutions  $J$  such that  $J\mathcal{T}J = \mathcal{T}^{-1}$  always exist, but there is no obvious candidate among them to represent time reversal and draw conclusions from it.

In the context of discrete variables the star-shaped interaction pattern is implemented as follows. We may naively assume that the energy conserved by the dynamics  $\mathcal{T}$  is simply  $E_{\text{dec}}$ , as if there were no energy for the interactions between the small part and the large ones, but  $\mathcal{T}$  must reflect the fact that the large parts interact only indirectly: there is an internal interaction energy  $\mathcal{E}(\mathcal{C})$  for every configuration  $\mathcal{C}$  of the small part and each change in the small part can be associated to an elementary energy exchange with one of the large parts. If the small part can jump from configuration  $\mathcal{C}$  to configuration  $\mathcal{C}'$  in a single time step by exchanging energy with the large part  $a$ , we use the notation  $\mathcal{C}' = \mathbb{F}_a \mathcal{C}$ . In this configuration jump the energy  $E_{\text{dec}}$  of the global system is conserved and the energy of the large part  $a$  is changed from  $E_a$  to  $E'_a$  according to the conservation law

$$E'_a - E_a = \begin{cases} -[\mathcal{E}(\mathcal{C}') - \mathcal{E}(\mathcal{C})] & \text{if } \mathcal{C}' = \mathbb{F}_a \mathcal{C} \\ 0 & \text{otherwise,} \end{cases} \quad (2.2)$$

while the energies of the other large parts are unchanged. Apart from these energy exchange constraints and from ergodicity, the deterministic dynamics  $\mathcal{T}$  is supposed to obey some other natural physical constraints which will be specified later.

### 2.1.3 Constraints from ergodicity and interaction pattern upon coarse-grained evolution

We start from the familiar observation that keeping track of what happens in detail in the large parts is out of our abilities, and often not very interesting anyway. Our ultimate interest is in the evolution of the configuration  $\mathcal{C}$  of the small part in an appropriate limit. As an intermediate step, we keep track also of the energies  $E_a$ 's in the large parts, but not of the detailed configurations in the large parts.

The coarse graining, which keeps track only of the time evolution of the configuration  $\mathcal{C}$  of the small part and the energies  $E_a$ 's of the large parts, is defined as follows. To each microscopic configuration  $\xi$  of the full system we can associate the corresponding configuration  $\mathcal{C} = \mathcal{C}(\xi)$  of the small part and the corresponding energy  $E_a = E_a(\xi)$  carried by part  $a$ . To simplify the notation, we let  $\underline{E}$  denote the collection of  $E_a$ 's, so  $\underline{E}$  is a vector with as many coordinates as there are large parts.

As shown in previous subsection, if the initial value of the energy  $E_{\text{dec}}$  is equal to  $E$ , then, over a period equal to  $N = \Omega_{\text{dec}}(E)$  in time step units, the trajectory of  $\xi$  under any microscopic ergodic dynamics  $\mathcal{T}$  corresponds to a cyclic permutation of all the microscopic configurations in the energy level  $E_{\text{dec}} = E$ . Thus ergodicity entails that at the coarse grained level, if  $N_{(\mathcal{C}, \underline{E})}$  denotes the occurrence number of  $(\mathcal{C}, \underline{E})$  during the period of  $N$  time steps,  $N_{(\mathcal{C}, \underline{E})}$  is nothing but  $\Omega_{\text{dec}}(\mathcal{C}, \underline{E})$ , the number of microscopic configurations of the full system when the small part is in configuration  $\mathcal{C}$  and the large parts have energies  $E_a$ 's in the energy level  $E = E_{\text{dec}}$ , namely

$$N(\mathcal{C}, \underline{E}) = \Omega_{\text{dec}}(\mathcal{C}, \underline{E}) \quad (2.3)$$

with

$$E = E_{\text{dec}} \equiv \mathcal{E}(\mathcal{C}) + \sum_a E_a. \quad (2.4)$$

In the following we fix the value of  $E$  and  $N$  is called the period of the dynamics, while the energy constraint  $E = \mathcal{E}(\mathcal{C}) + \sum_a E_a$  is often implicit in the notations.

Another crucial point is that, since any specific dynamics  $\mathcal{T}$  under consideration respects both the conservation of  $E_{\text{dec}}$  and the interaction pattern specified at the end of subsection 2.1.2, the number of jumps from  $(\mathcal{C}, \underline{E})$  to  $(\mathcal{C}', \underline{E}')$  over the period of  $N$  time steps, denoted by  $N_{(\mathcal{C}, \underline{E}), (\mathcal{C}', \underline{E}')}$ , is equal to the number of the reversed jumps from  $(\mathcal{C}', \underline{E}')$  to  $(\mathcal{C}, \underline{E})$  during the same time interval, namely

$$N_{(\mathcal{C}, \underline{E}), (\mathcal{C}', \underline{E}')} = N_{(\mathcal{C}', \underline{E}'), (\mathcal{C}, \underline{E})}. \quad (2.5)$$

In Appendix A we give graph-theoretic conditions, not related to ergodicity, that ensure this property, and show that they are fulfilled in one relevant example, as a consequence of the star-shaped interaction pattern.

## 2.2 Markovian approximation for the mesoscopic dynamics

### 2.2.1 Definition of a Markovian approximation for the mesoscopic dynamics

As already noticed, over the period of  $N = \Omega_{\text{dec}}(E)$  time steps a trajectory under any microscopic ergodic dynamics  $\mathcal{T}$  in the energy level  $E_{\text{dec}} = E$  corresponds to a cyclic permutation of the  $N$  microscopic configurations  $\xi$ 's in the energy level. Therefore, if the configuration at some initial time is denoted by  $\xi_1$ , then the trajectory is represented by the sequence  $\omega = \xi_1 \xi_2 \cdots \xi_N$  where  $\xi_{i+1} = \mathcal{T}\xi_i$  with  $\xi_{N+1} = \xi_1$ . By the coarse-grained procedure which retains only the mesoscopic variable  $x \equiv (\mathcal{C}, \underline{E})$ , the succession of distinct microscopic configurations  $\omega$  is replaced by  $w = x_1 x_2 \cdots x_N$  where  $x_i = x(\xi_i) = (\mathcal{C}(\xi_i), \underline{E}(\xi_i))$ . In  $w$  various  $x_i$ 's take the same value, and a so-called transition corresponds to the case  $x_{i+1} \neq x_i$ , namely the case where the configuration  $\mathcal{C}$  of the small system is changed in the jump of the microscopic configuration of the full system from  $\xi_i$  to  $\xi_{i+1} = \mathcal{T}\xi_i$ .

Since the large parts involve many degrees of freedom, the number of times some given value  $x = (\mathcal{C}, \underline{E})$  appears in the coarse-grained sequence  $w$  is huge, and even if  $\omega$  is given by the deterministic rule  $\xi_{i+1} = \mathcal{T}\xi_i$ , there is no such rule to describe the sequence  $w$ . Moreover the microscopic configuration at the initial time,  $\xi_1$ , is not known so that the coarse grained sequence which actually appears in the course of time is in fact a sequence deduced from  $w$  by a translation of all indices.

As explained in Appendix B, one may associate to the sequence  $w$  a (discrete time) Markov chain such that the mean occurrence frequencies of the patterns  $x$  and  $xx'$  in a stationary stochastic sample are equal to the corresponding values,  $N_x/N$  and  $N_{xx'}/N$ , in the sequence  $w$  determined by the dynamics  $\mathcal{T}$  (up to a translation of all indices corresponding to a different value of the initial microscopic configuration). Whether this Markovian effective description is accurate depends on several things: the choice of  $\mathcal{T}$ , the kind of statistical properties of  $w$  one wants to check, etc.

We may also argue that a continuous time description is enough if we restrict our attention to microscopic dynamical maps  $\mathcal{T}$ 's such that transitions, namely the patterns  $xx'$  with  $x' \neq x$ , are rare and of comparable mean occurrence frequencies over the period of  $N$  time steps. In other words, most of the steps in the dynamics amount to reshuffle the configurations of the large parts without changing their energies, leaving the configuration of the small part untouched. The latter physical constraint and the hypothesis of the validity of the Markovian approximation select a particular class of dynamics  $\mathcal{T}$ .

With these assumptions, we associate to the sequence  $w$  of coarse grained variables  $(\mathcal{C}, \underline{E})$  a Markov process whose stationary measure shares some of the statistical properties of  $w$ , namely the values of the mean occurrence frequencies of length 1 and length 2 patterns. The transition rate from  $(\mathcal{C}, \underline{E})$  to  $(\mathcal{C}', \underline{E}')$  with  $(\mathcal{C}, \underline{E}) \neq (\mathcal{C}', \underline{E}')$  in the approximated Markov process is given by (B.4), where we just have to make the substitutions  $N_x = N_{(\mathcal{C}, \underline{E})}$  and  $N_{xx'} = N_{(\mathcal{C}, \underline{E}), (\mathcal{C}', \underline{E}')}$ , while the corresponding stationary distribution is given by (B.2),  $P_{\text{st}}^W(\mathcal{C}, \underline{E}) = N_{(\mathcal{C}, \underline{E})}/N$  with  $N = \sum_{(\mathcal{C}, \underline{E})} N_{(\mathcal{C}, \underline{E})}$ .

### 2.2.2 Microcanonical detailed balance and other properties

By virtue of the ergodicity property (2.3) at the coarse-grained level, the transition rate in the approximated Markov process reads

$$W(\mathcal{C}', \underline{E}' \leftarrow \mathcal{C}, \underline{E}) = \frac{N_{(\mathcal{C}, \underline{E}), (\mathcal{C}', \underline{E}')}}{\tau \Omega_{\text{dec}}(\mathcal{C}, \underline{E})} \quad \text{for } (\mathcal{C}, \underline{E}) \neq (\mathcal{C}', \underline{E}'), \quad (2.6)$$

where  $\tau$  is a time scale such that  $W(\mathcal{C}', \underline{E}' \leftarrow \mathcal{C}, \underline{E})$  is of order unity. Meanwhile, by virtue of the ergodicity properties (2.1) and (2.3), the corresponding stationary distribution is merely the microcanonical distribution

$$P_{\text{st}}^W(\mathcal{C}, \underline{E}) = \frac{\Omega_{\text{dec}}(\mathcal{C}, \underline{E})}{\Omega_{\text{dec}}(E)} \equiv P_{\text{mc}}(\mathcal{C}, \underline{E}). \quad (2.7)$$

Ergodicity also entails that, since all microscopic configurations  $\xi$ 's in the energy level appear in the sequence  $\omega$ , all possible values of  $x$  also appear in the coarse-grained sequence  $w$  : so any mesoscopic state  $(\mathcal{C}, \underline{E})$  can be reached from any other mesoscopic state  $(\mathcal{C}', \underline{E}')$  by a succession of elementary transitions, even if they are not involved in an elementary transition (i.e. if  $N_{(\mathcal{C}, \underline{E}), (\mathcal{C}', \underline{E}')} = 0$ ); in other words the graph associated to the transition rates  $W(\mathcal{C}', \underline{E}' \leftarrow \mathcal{C}, \underline{E})$  is connected, or, equivalently, the Markov matrix defined from the transition rates (see definition below in (2.15)) is irreducible.

The constraint (2.5) imposed by the interaction pattern upon the coarse-grained evolution over a period of  $N$  time steps entails that the transition rates of the approximated Markov process defined in (2.6) obey two properties. First,

$$W(\mathcal{C}', \underline{E}' \leftarrow \mathcal{C}, \underline{E}) \quad \text{and} \quad W(\mathcal{C}, \underline{E} \leftarrow \mathcal{C}', \underline{E}') \quad (2.8)$$

are either both  $= 0$  or both  $\neq 0$ . This property may be called microreversibility. Second, if the transition rates do not vanish, they obey the equality

$$\frac{W(\mathcal{C}', \underline{E}' \leftarrow \mathcal{C}, \underline{E})}{W(\mathcal{C}, \underline{E} \leftarrow \mathcal{C}', \underline{E}')} = \frac{\Omega_{\text{dec}}(\mathcal{C}', \underline{E}')}{\Omega_{\text{dec}}(\mathcal{C}, \underline{E})} = \frac{P_{\text{mc}}(\mathcal{C}', \underline{E}')}{P_{\text{mc}}(\mathcal{C}, \underline{E})}. \quad (2.9)$$

Observe that the arbitrary time scale  $\tau$  has disappeared in this equation. The equality between the ratio of transition rates and the ratio of probabilities in the corresponding stationary distribution is the so-called detailed balance relation. Here the stationary distribution is that of the microcanonical ensemble, and we will refer to relation (2.9) as the microcanonical detailed balance.

The microcanonical detailed balance (2.9) for the transition rates  $W(\mathcal{C}', \underline{E}' \leftarrow \mathcal{C}, \underline{E})$  can be derived in another context where the discrete dynamical variables are replaced by continuous mesoscopic observables evolving under a Markov process which arises from an underlying microscopic dynamics whose Hamiltonian is an even function of momenta (see the derivation in Appendix C). In the latter context the fact that the stationary measure is the microcanonical distribution is enforced by the invariance of the Liouville measure in phase space under the Hamiltonian evolution (see the derivation of (C.5)); the microcanonical detailed balance mainly arises from the invariance under time reversal of the trajectories in phase space (see (C.7)). In our argument for discrete variables, the fact that the stationary measure is the microcanonical distribution arises from the ergodicity of the microscopic map  $\mathcal{T}$  (see (2.1) and (2.3)); the microcanonical detailed balance emerges from the equality between the frequencies of a given transition and the reversed one over a period of the microscopic map determined by the total energy, equality which is enforced by the star-shaped interaction pattern (see (2.5)).

### 2.2.3 Further consequence of the interaction pattern

In the interaction pattern large parts do not interact directly with one another and the energy of a single large part is changed in a given transition. We let  $\Omega_a(E_a)$  be the number of configurations in large part  $a$  with energy  $E_a$  when it is isolated. Then  $\Omega_{\text{dec}}(\mathcal{C}, \underline{E}) = \prod_a \Omega_a(E_a)$ , so if  $(\mathcal{C}', \underline{E}')$  is obtained from  $(\mathcal{C}, \underline{E})$  by an energy exchange with bath  $a$  which makes  $\mathcal{C}$  jump to  $\mathcal{C}'$  we have the result, with the notation introduced in (2.2),

$$\text{if } \mathcal{C}' = \mathbb{F}_a \mathcal{C} \quad \frac{\Omega_{\text{dec}}(\mathcal{C}', \underline{E}')}{\Omega_{\text{dec}}(\mathcal{C}, \underline{E})} = \frac{\prod_b \Omega_b(E'_b)}{\prod_b \Omega_b(E_b)} = \frac{\Omega_a(E'_a)}{\Omega_a(E_a)}. \quad (2.10)$$

We have used that  $E'_b = E_b - \delta_{a,b} [\mathcal{E}(\mathcal{C}') - \mathcal{E}(\mathcal{C})]$ , so that for  $b \neq a$  the multiplicity factors are unchanged in the transition.

The latter ratio of microstate numbers can be expressed in terms of the Boltzmann entropies when each part  $a$  is isolated. The dimensionless Boltzmann entropy<sup>2</sup>  $S_a^B(E_a)$  for the isolated part  $a$  when its energy is equal to  $E_a$  is defined by  $\Omega_a(E_a) = \exp S_a^B(E_a)$ . With these notations, if the transition rate  $W(\mathcal{C}', \underline{E}' \xleftarrow{\mathbb{F}_a} \mathcal{C}, \underline{E})$  where  $\mathcal{C}' = \mathbb{F}_a \mathcal{C}$  is nonzero, then the transition rate for the

<sup>2</sup>The Boltzmann constant is set equal to 1.

reversed jump  $W(\mathcal{C}, \underline{E} \xrightarrow{\mathbb{F}_a} \mathcal{C}', \underline{E}')$  where  $\mathcal{C} = \mathbb{F}_a \mathcal{C}'$  is also non zero (see (2.8)) and, by virtue of (2.10) the relation (2.9) is reduced to

$$\frac{W(\mathcal{C}', \underline{E} \xrightarrow{\mathbb{F}_a} \mathcal{C}, \underline{E})}{W(\mathcal{C}, \underline{E} \xrightarrow{\mathbb{F}_a} \mathcal{C}', \underline{E}')} = \frac{\Omega_a(E'_a)}{\Omega_a(E_a)} \equiv e^{S_a^B(E'_a) - S_a^B(E_a)}, \quad (2.11)$$

The latter formula is the first important stage of the argument. It has already been used in the literature, for instance in Ref.[40].

We stress that the present argument does not involve any kind of underlying time reversal. Here the time reversal symmetry arises only at the statistical level of description represented by the Markov evolution ruled by the transition rates.

Notice also that, as only certain ratios are fixed, different ergodic deterministic microscopic dynamics can lead to very different transition rates, a remnant of the fact that the coupling between a large part and the small part can take any value a priori.

Formula (2.11) is also a clue to understand a contrario what kind of physical input is needed for the homogeneous Markov approximation to be valid. Indeed, why didn't we do the homogeneous Markov approximation directly on the small part ? We could certainly imagine dynamics making this a valid choice. However, it is in general incompatible with the pattern of interactions (see the end of subsection 2.1.2) which is the basis of our argument. For instance, if in the coarse graining procedure we had retained only the configurations  $\mathcal{C}$ 's of the small part, then the corresponding graph introduced in Appendix A would have been a cycle instead of a tree in the case of a small part made of two spins (see paper II [45]), and the crucial property (2.5) would have been lost : over a period of  $N$  time steps of the microscopic dynamics  $N_{\mathcal{C}, \mathcal{C}'} \neq N_{\mathcal{C}', \mathcal{C}}$ . In fact, we may expect, or impose on physical grounds, that  $\Omega_a$  will be exponentially large in the size of the large part  $a$  (i.e. its number of degrees of freedom  $\mathcal{N}_a$ ), so that even the ratio  $\Omega_a(E'_a)/\Omega_a(E_a) = \Omega_a(E_a - [\mathcal{E}(\mathcal{C}') - \mathcal{E}(\mathcal{C})])/\Omega_a(E_a)$  will vary significantly over the trajectory, meaning that transition probabilities for the small part only cannot be taken to be constant along the trajectory: the energies of the large parts are relevant variables.

## 2.3 Transient regime when large parts are described in the thermodynamic limit

### 2.3.1 Large parts in the thermodynamic limit

We now assume that the large parts are large enough that they are accurately described by a thermodynamic limit, which we take at the most naive level. To recall what we mean by that, we concentrate on one large part for a while, and suppress the index used to label it. Suppose this large part has  $\mathcal{N}$  degrees of freedom, and suppose that energies are close to an energy  $E$  for which the entropy is  $S^B(E)$ . That the thermodynamic limit exists means that if one lets  $\mathcal{N} \rightarrow +\infty$  while the ratio  $E/\mathcal{N}$  goes to a finite limit  $\epsilon$ , there is a differentiable function  $s^B(\epsilon)$  such that the ratio  $S^B(E)/\mathcal{N}$  goes to  $s^B(\epsilon)$ . The quantity

$$\frac{ds^B}{d\epsilon} \equiv \beta \quad (2.12)$$

is nothing but the inverse temperature. In that case, as long as  $\Delta E \ll \mathcal{N}\delta E$  (where  $\delta E$  is some microscopic energy scale)  $S^B(E + \Delta E) - S^B(E) \rightarrow \beta\Delta E$  when  $\mathcal{N} \rightarrow +\infty$ . For  $\mathcal{N}$  large enough, the relation  $S^B(E + \Delta E) - S^B(E) \sim \beta\Delta E$  is a good approximation.

### 2.3.2 Transient regime and modified detailed balance (MDB)

Notice that when transitions occur, which, by the definition of  $\tau$  in (2.6), happens typically once on the macroscopic time scale, the changes in the energies of the large parts are finite, so that over long windows of time evolution, involving many changes in the small part, the relation

$$S_a^B(E'_a) - S_a^B(E_a) \sim \beta_a [E'_a - E_a] \quad (2.13)$$

is not spoiled, where  $E'_a$  and  $E_a$  are the energies in large part  $a$  at any moment within the window.

In fact the larger the large parts, the longer the time window for which (2.13) remains valid. The relation between the sizes  $\mathcal{N}_a$ 's of the large parts and of the length of the time window depends on the details of  $\mathcal{T}$  (which still has to fulfill the imposed physical conditions). This relation also depends on the values of the energy per degree of freedom in every large part,  $E_a/\mathcal{N}_a$ , which are essentially constant in such a window.

Because of the ergodicity hypothesis, the largest window (of size comparable to the period of  $\mathcal{T}$  to logarithmic precision), has the property that the energies  $E_a$ 's in the large parts will be such that all  $\beta_a$ 's are close to each other and the system will be at equilibrium. Indeed, inside the largest time window, the system remains in the region of the energy level where the energies  $E_a$ 's are the most probable, and in the thermodynamic limit the most probable values for the  $E_a$ 's in the microcanonical ensemble are the values  $E_a^*$ 's which maximize the product  $\prod_a \Omega_a(E_a)$  under the constraint  $E = \sum_a E_a$  (since the system energies are negligible with respect of those of the large parts). The latter maximization condition is equivalent to the equalities  $ds^B/d\epsilon_a(\epsilon_a^*) = ds^B/d\epsilon_b(\epsilon_b^*)$  for all pairs of large parts.

However, if the system starts in a configuration such that the  $\beta_a$ 's are distinct, the time window over which  $E_a/\mathcal{N}_a$  and  $\beta_a$  are constant (to a good approximation) will be short with respect to the period of the microscopic dynamics, but long enough that (2.13) still holds for a long time interval. Then by putting together the information on the ratio of transition rates in terms of Boltzmann entropies (2.11), the transient regime approximation (2.13) and the energy conservation (2.2), we get

$$\frac{W(\mathcal{C}', \underline{E}' \xleftarrow{\mathbb{F}_a} \mathcal{C}, \underline{E})}{W(\mathcal{C}, \underline{E} \xleftarrow{\mathbb{F}_a} \mathcal{C}', \underline{E}')} \sim e^{-\beta_a [\mathcal{E}(\mathcal{C}') - \mathcal{E}(\mathcal{C})]}. \quad (2.14)$$

Now the right-hand side depends only on the configurations of the small system, and the parameters  $\beta_a$ 's are constants. Letting the large parts get larger and larger while adjusting the physical properties adequately, we can ensure that the time over which (2.14) remains valid gets longer and longer, so in the thermodynamic limit for the large parts, the transient regime lasts forever. This situation is our main interest in what follows.

The transition rates in the transient regime correspond to a Markov matrix  $\mathbb{M}$  defined by

$$(\mathcal{C}' | \mathbb{M} | \mathcal{C}) = \begin{cases} (\mathcal{C}' | \mathbb{W} | \mathcal{C}) & \text{if } \mathcal{C}' \neq \mathcal{C} \\ -\sum_{\mathcal{C}'} (\mathcal{C}' | \mathbb{W} | \mathcal{C}) & \text{if } \mathcal{C}' = \mathcal{C}. \end{cases} \quad (2.15)$$

As well as the transition rates  $W(\mathcal{C}', \underline{E}' \xleftarrow{\mathbb{F}_a} \mathcal{C}, \underline{E})$  they must satisfy the three consequences derived from the properties of the underlying microscopic deterministic dynamics pointed out in subsubsection 2.1.2, namely ergodicity, energy conservation and specific interaction pattern. First, as shown in subsubsection 2.2.2, the Markov matrix is irreducible, or in other words the graph associated with the transition rates is connected (see (1.1)). Second, from (2.8) the transition rates must obey the microscopic reversibility condition (1.2) for any couple of configurations  $(\mathcal{C}, \mathcal{C}')$ . Third, from (2.14) one gets a constraint obeyed by the ratio of transition rates in the transient regime,

$$\text{for } \mathcal{C}' = \mathbb{F}_a \mathcal{C} \quad \frac{(\mathcal{C}' | \mathbb{W} | \mathcal{C})}{(\mathcal{C} | \mathbb{W} | \mathcal{C}')} = e^{-\beta_a [\mathcal{E}(\mathcal{C}') - \mathcal{E}(\mathcal{C})]}. \quad (2.16)$$

The latter relation is the so-called modified detailed balance (MDB), which is also referred to in the literature as the “generalized detailed balance”.

We stress that, by selecting a time window while taking the thermodynamic limit for the large parts, the microcanonical detailed balance (2.9) is replaced by the modified detailed balance (2.11), except in the case of the largest time window where all  $\beta_a$ 's are equal. In the latter case, the microcanonical detailed balance (2.9) is replaced by the canonical detailed balance and the statistical time reversal symmetry is preserved. Indeed, the equilibrium thermodynamic regime is reached either if we start from a situation in which  $\prod_a \Omega_a(E_a)$  is close to its maximum along the trajectory in the energy level  $E_{\text{dec}} = E$ , or if we wait long enough so that  $\prod_a \Omega_a(E_a)$

becomes close to this maximum. As recalled above, this is true for most of the period of the microscopic dynamics, but reaching this situation may however take a huge number of time steps if the starting point was far from the maximum. By an argument similar to that used in the derivation of (2.13), when  $\prod_a \Omega_a(E_a)$  is closed to its maximum and the large parts are considered in the thermodynamic limit, all  $\beta_a$ 's are equal to the same value  $\beta$  and the relative weight of two configurations in the microcanonical ensemble,  $P_{\text{mc}}(\mathcal{C}', \underline{E}')/P_{\text{mc}}(\mathcal{C}, \underline{E})$  given by (2.7), is shown to tend to  $\exp(-\beta[\mathcal{E}(\mathcal{C}') - \mathcal{E}(\mathcal{C})])$ . Then the equilibrium microcanonical distribution  $P_{\text{mc}}(\mathcal{C}, \underline{E})$  tends to the canonical distribution

$$P_{\text{can}}^\beta(\mathcal{C}) \equiv \frac{e^{-\beta\mathcal{E}(\mathcal{C})}}{Z(\beta)}, \quad (2.17)$$

where  $Z(\beta)$  is the canonical partition function at the inverse temperature  $\beta$ . Meanwhile, the detailed balance relation (2.9) in the microcanonical equilibrium ensemble for the transition rates  $W(\mathcal{C}', \underline{E}' \leftarrow \mathcal{C}, \underline{E})$  becomes a detailed balance relation in the canonical ensemble at the inverse temperature  $\beta$  of the whole system for the transition rates  $(\mathcal{C}'|\mathbb{W}|\mathcal{C})$ , namely

$$\frac{(\mathcal{C}'|\mathbb{W}|\mathcal{C})}{(\mathcal{C}|\mathbb{W}|\mathcal{C}')} = e^{-\beta[\mathcal{E}(\mathcal{C}') - \mathcal{E}(\mathcal{C})]} = \frac{P_{\text{can}}^\beta(\mathcal{C}')}{P_{\text{can}}^\beta(\mathcal{C})}. \quad (2.18)$$

The modified detailed balance (2.16), valid in transient regimes, differs from the latter detailed balance in the canonical ensemble by two features : the various  $\beta_a$ 's of the distinct large parts appears in place of the common equilibrium inverse temperature  $\beta$ , and the stationary distribution for the transition rates is not known a priori.

We conclude this discussion with the following remarks. We have not tried to exhibit explicit physical descriptions of the large parts, or explicit formulæ for the dynamical map  $\mathcal{T}$ . Though it is not too difficult to give examples for fixed sizes of the large parts, it is harder to get a family of such descriptions sharing identical physical properties for varying large part sizes, a feature which is crucial to really make sense of the limits we took blindly in our derivation. It is certainly doable, but cumbersome, and we have not tried to pursue this idea. Let us note also that in principle, taking large parts of increasing sizes can be used to enhance the validity of the approximation of the (discrete-time) Markov chain by a (continuous-time) Markov process. As the physics of the continuous time limit does not seem to be related to the physics of convergence towards a heat bath description we have preferred to keep the discussion separate, taking a continuous-time description as starting point.

### 2.3.3 Expression of MDB in terms of exchange entropy variation

Observe that though we have given no detailed analysis of the dynamics or the statistical properties of the large parts, their influence on the effective Markov dynamics of the small system enters only through the inverse temperatures  $\beta_a$  defined in (2.12). So we can consistently assume that each large part becomes a thermal bath with its own temperature. The leading term in  $S_a^B(E'_a) - S_a^B(E_a)$  is the variation  $\delta S_a^{TH}(\mathcal{C}' \leftarrow \mathcal{C})$  of the thermodynamic entropy of bath  $a$  when it flips the system from configuration  $\mathcal{C}$  to configuration  $\mathcal{C}'$ ,

$$\delta S_a^{TH}(\mathcal{C}' \leftarrow \mathcal{C}) = \begin{cases} \beta_a [E'_a - E_a] & \text{if } \mathcal{C}' = \mathbb{F}_a \mathcal{C} \\ 0 & \text{otherwise.} \end{cases} \quad (2.19)$$

Then we have an idealized description of a thermal contact between to heat baths. This is the situation on which we concentrate in this paper.

Since  $\mathcal{E}(\mathbb{F}_a \mathcal{C}) - \mathcal{E}(\mathcal{C})$  corresponds to a variation in the energy level populations of the two-spin system, it corresponds to the heat  $\delta q_a(\mathcal{C}' \leftarrow \mathcal{C})$  received by the two-spin system from the part  $a$  when it evolves from configuration  $\mathcal{C}$  to configuration  $\mathcal{C}' = \mathbb{F}_a \mathcal{C}$ ,

$$\begin{cases} \delta q_a(\mathcal{C}' \leftarrow \mathcal{C}) = \mathcal{E}(\mathcal{C}') - \mathcal{E}(\mathcal{C}) & \text{if } \mathcal{C}' = \mathbb{F}_a \mathcal{C} \\ \delta q_a(\mathcal{C}' \leftarrow \mathcal{C}) = 0 & \text{otherwise.} \end{cases} \quad (2.20)$$



According to the energy conservation relation (2.2) the variation  $\delta S_a^{TH}(\mathcal{C}' \leftarrow \mathcal{C})$  of the thermodynamic entropy of the bath  $a$  when it flips the system from configuration  $\mathcal{C}$  to configuration  $\mathcal{C}'$ , defined in (2.19), reads

$$\delta S_a^{TH}(\mathcal{C}' \leftarrow \mathcal{C}) = -\beta_a \delta q_a(\mathcal{C}' \leftarrow \mathcal{C}). \quad (2.21)$$

Let us introduce the exchange entropy variation<sup>3</sup> of the system  $\delta_{\text{exch}} S(\mathcal{C}' \leftarrow \mathcal{C})$  which is associated with the heat exchanges with the two thermostats when the two-spin system goes from configuration  $\mathcal{C}$  to configuration  $\mathcal{C}'$ . Thanks to the definition (2.19)

$$\delta_{\text{exch}} S(\mathcal{C}' \leftarrow \mathcal{C}) \equiv -\sum_a \delta S_a^{TH}(\mathcal{C}' \leftarrow \mathcal{C}), \quad (2.22)$$

namely

$$\delta_{\text{exch}} S(\mathcal{C}' \leftarrow \mathcal{C}) \equiv \sum_a \beta_a \delta q_a(\mathcal{C}' \leftarrow \mathcal{C}). \quad (2.23)$$

Then the modified detailed balance (2.16) can be rewritten in a form which does not involve explicitly the heat bath responsible for the transition from  $\mathcal{C}$  to  $\mathcal{C}'$ ,

$$\frac{(\mathcal{C}'|\mathbb{W}|\mathcal{C})}{(\mathcal{C}|\mathbb{W}|\mathcal{C}')} = e^{-\delta_{\text{exch}} S(\mathcal{C}' \leftarrow \mathcal{C})}. \quad (2.24)$$

### 3 Master equation, exchange entropy flow and various entropy variations

#### 3.1 Evolution of the probability distribution without MDB (known results)

In this subsection we recall previously known results which are important milestones to our original results and which are consequences of the first two properties (1.1) and (1.2), among the three mesoscopic conditions derived from the ergodicity of the underlying conservative deterministic dynamics.

The starting point is that the evolution of the probability  $P(\mathcal{C}; t)$  that the system is in configuration  $\mathcal{C}$  at time  $t$  is ruled by the master equation

$$\frac{dP(\mathcal{C}; t)}{dt} = \sum_{\mathcal{C}' \neq \mathcal{C}} (\mathcal{C}|\mathbb{W}|\mathcal{C}') P(\mathcal{C}'; t) - \sum_{\mathcal{C}' \neq \mathcal{C}} (\mathcal{C}'|\mathbb{W}|\mathcal{C}) P(\mathcal{C}; t) = \sum_{\mathcal{C}'} (\mathcal{C}|\mathbb{M}|\mathcal{C}') P(\mathcal{C}'; t), \quad (3.1)$$

where  $\mathbb{M}$  is defined in (2.15). Since  $P(\mathcal{C}; t)$  is to be interpreted as a probability distribution, it has to satisfy the positivity and normalization conditions,

$$\forall \mathcal{C} \quad P(\mathcal{C}; t) \geq 0 \quad \text{and} \quad \sum_{\mathcal{C}} P(\mathcal{C}; t) = 1. \quad (3.2)$$

##### 3.1.1 Generic properties

According to the theory of systems of ordinary differential equations with constant coefficients, the solution  $P(t)$  of the equation (3.1) exists and is unique for a given initial function  $P(t_0)$ . The Markov matrix  $\mathbb{M}$  obeys the property  $\sum_{\mathcal{C}, \mathcal{C}'} (\mathcal{C}|\mathbb{M}|\mathcal{C}') P(\mathcal{C}'; t) = 0$  for any  $P$  and this ensures that the normalization constraint is preserved under the time evolution,

$$\sum_{\mathcal{C}} P(\mathcal{C}; t = 0) = 1 \quad \Rightarrow \quad \forall t > 0 \quad \sum_{\mathcal{C}} P(\mathcal{C}; t) = 1. \quad (3.3)$$

The above property of  $\mathbb{M}$  also ensures the existence of at least one stationary solution, but there is no argument for every stationary solution to obey the positivity constraint in (3.2) without further assumptions.

---

<sup>3</sup> Exchange entropy variation is an abbreviation for variation of entropy due to heat exchanges with a bath.

### 3.1.2 Properties arising from the irreducibility condition

When the transition rates satisfy the irreducibility condition (1.1), and if positivity and normalization (3.2) are satisfied at the initial time, the solution of the master equation (3.1) not only meets these conditions for being a probability distribution at any subsequent time, but it even has the more stringent property that any configuration has a strictly non-vanishing weight,

$$\forall t > 0 \quad \forall \mathcal{C} \quad P(\mathcal{C}; t) > 0. \quad (3.4)$$

This result can be derived in the framework of the theory of ordinary differential equations with constant coefficients [52].

Moreover the irreducibility condition (1.1) allows one to build at least formally a stationary solution which fulfills the conditions (3.2) for being a probability (with the even more stringent property (3.4)). This solution is obtained in the framework of graph theory by using the network representation of the master equation; the corresponding expression is called Kirchhoff's theorem in Ref.[52].

The irreducibility condition (1.1) also entails that the stationary solution of the master equation is unique, and therefore coincides with the expression given by Kirchhoff's theorem. The uniqueness of the stationary solution can be derived either in the framework of matrix theory by using the Perron-Frobenius theorem (see for instance [47]) or in the framework of the theory of ordinary differential equations with constant coefficients [52].

The latter derivation uses the stability criterion introduced by Schögl [53]. As noticed by Schögl a good candidate for the Liapunov function involved in a stability criterion is the relative entropy introduced by Kullback and Leibler [54] in the context of information theory, namely

$$S_{\text{rel}}[P(t)|P_{\text{st}}] \equiv \sum_{\mathcal{C}} P(\mathcal{C}; t) \ln \frac{P(\mathcal{C}; t)}{P_{\text{st}}(\mathcal{C})}. \quad (3.5)$$

The relative entropy is well-defined at any time according to (3.4). (In the generic case  $S_{\text{rel}}[P|P_0]$  is well defined only if  $P_0(\mathcal{C}) = 0$  implies  $P(\mathcal{C}) = 0$ .) The definition of  $S_{\text{rel}}[P|P_0]$ , with any given  $P_0$ , entails that  $S_{\text{rel}}[P|P_0]$  is positive for any  $P$  and vanishes only when  $P$  is equal to  $P_0$  (by virtue of the inequality  $\ln x < x - 1$  if  $x > 0$  and  $x \neq 1$ ). Therefore

$$S_{\text{rel}}[P|P_{\text{st}}] > 0 \quad \text{if } P \neq P_{\text{st}} \quad \text{and} \quad S_{\text{rel}}[P_{\text{st}}|P_{\text{st}}] = 0. \quad (3.6)$$

The definition also ensures that  $S_{\text{rel}}[P|P_0]$  is convex (i.e. concave upward) for any  $P$ , so that

$$\delta^{(2)} S_{\text{rel}}[P|P_{\text{st}}] > 0 \quad \text{for any } P, \quad (3.7)$$

with

$$\delta^{(2)} f[P] \equiv \frac{1}{2} \sum_{\mathcal{C}, \mathcal{C}'} \left. \frac{\partial^2 f}{\partial \tilde{P}(\mathcal{C}) \partial \tilde{P}(\mathcal{C}')} \right|_{\tilde{P}=P} \delta P(\mathcal{C}) \delta P(\mathcal{C}') \quad (3.8)$$

and  $\delta P(\mathcal{C}) \equiv \tilde{P}(\mathcal{C}) - P(\mathcal{C})$ . Moreover, because of the structure of the master equation (3.1) combined with the properties  $\ln x \leq x - 1$  for any  $x > 0$  and  $\sum_{\mathcal{C}'} (\mathcal{C}|\mathbb{M}|\mathcal{C}') P_{\text{st}}(\mathcal{C}') = 0$ , the time derivative of  $S_{\text{rel}}[P(t)|P_{\text{st}}]$  is negative at any time [52],

$$\frac{dS_{\text{rel}}[P(t)|P_{\text{st}}]}{dt} \leq 0. \quad (3.9)$$

The properties (3.6) and (3.9) define a Liapunov function and ensure that for any initial distribution  $P(t_0)$  in the vicinity of  $P_{\text{st}}$   $\lim_{t \rightarrow +\infty} P(t) = P_{\text{st}}$  (because the property (3.6) ensures that  $\delta^{(2)} S_{\text{rel}}[P|P_{\text{st}}] > 0$  for any  $P$  in the vicinity of  $P_{\text{st}}$ ). With the extra property (3.7) one can apply the stability theorem by Schögl which states that  $\lim_{t \rightarrow +\infty} P(t) = P_{\text{st}}$  for any initial distribution  $P(t_0)$ . The interpretation given by Schögl of the stability condition  $dS_{\text{rel}}[P(t)|P_{\text{st}}]/dt \leq 0$  can be rephrased as follows. If the observer does not know more than that the system was initially in

some unknown state  $P_0$  of the stability region, then an unbiased estimate for the state  $P(t)$  at time  $t$  would be  $P_{\text{st}}$ . But if the observer knows  $P_0$  by measurements, then his excess knowledge at initial time is equal to  $S_{\text{rel}}[P_0|P_{\text{st}}]$ , and the property  $dS_{\text{rel}}[P(t)|P_{\text{st}}]/dt \leq 0$  reflects the fact that the spontaneous development of the states after the last observation can only go such that this knowledge does not increase.

## 3.2 Microscopic currents

### 3.2.1 Definitions

Our results in the framework of Markovian stochastic dynamics described by a master equation can be written in compact forms if we introduce the generic current  $j(\mathcal{C})$  for a microscopic variation when the system goes out of a given configuration  $\mathcal{C}$ . Such a current is associated either to the variation  $\mathcal{O}(\mathcal{C}') - \mathcal{O}(\mathcal{C})$  of a configuration observable  $\mathcal{O}(\mathcal{C})$  and then

$$j_{\mathcal{O}}(\mathcal{C}) \equiv \sum_{\mathcal{C}'} (\mathcal{C}'|\mathbb{W}|\mathcal{C}) [\mathcal{O}(\mathcal{C}') - \mathcal{O}(\mathcal{C})], \quad (3.10)$$

or more generally to some exchange quantity  $\delta K(\mathcal{C}' \leftarrow \mathcal{C})$  such as a heat amount or the variation of the reservoir entropies when the system jumps from configuration  $\mathcal{C}$  to  $\mathcal{C}'$ , and then

$$j_{\delta K}(\mathcal{C}) \equiv \sum_{\mathcal{C}'} (\mathcal{C}'|\mathbb{W}|\mathcal{C}) \delta K(\mathcal{C}' \leftarrow \mathcal{C}). \quad (3.11)$$

We include in the definition of an exchange quantity that it obeys the antisymmetry property  $\delta K(\mathcal{C}' \leftarrow \mathcal{C}) = -\delta K(\mathcal{C} \leftarrow \mathcal{C}')$ . We notice that the above definitions are valid even if the observable  $\mathcal{O}$  (or the variation  $\delta K(\mathcal{C}' \leftarrow \mathcal{C})$ ) depends explicitly on time.

The average of a current  $j(\mathcal{C})$  at time  $t$  is given by the generic formula for the mean value  $\langle \mathcal{O} \rangle_t$  of a configuration observable  $\mathcal{O}(\mathcal{C})$  at time  $t$ , namely

$$\langle \mathcal{O} \rangle_t \equiv \sum_{\mathcal{C}} \mathcal{O}(\mathcal{C}) P(\mathcal{C}; t). \quad (3.12)$$

According to the master equation (3.1) the time derivative of the mean value  $\langle \mathcal{O} \rangle_t$  of a configuration observable  $\mathcal{O}(\mathcal{C})$  which does not depend explicitly on time is equal to the mean value  $\langle j_{\mathcal{O}} \rangle_t$  of the associated current  $j_{\mathcal{O}}(\mathcal{C})$ ,

$$\frac{d\langle \mathcal{O} \rangle_t}{dt} = \langle j_{\mathcal{O}} \rangle_t, \quad (3.13)$$

where  $j_{\mathcal{O}}(\mathcal{C})$  is the current (3.10) of observable  $\mathcal{O}$  associated with the stochastic jumps going out of configuration  $\mathcal{C}$ .

### 3.2.2 Probabilistic interpretation of microscopic observable currents

To every observable i.e. to every real-valued function  $\mathcal{O}(\mathcal{C})$  on the configuration space, one can associate the random process  $\mathcal{O}(\mathcal{C}_t)$  where  $\mathcal{C}_t$  denotes the configuration of the system at time  $t$ . Note that  $P(\mathcal{C}; t)$  is nothing but  $P(\mathcal{C}_t = \mathcal{C})$ , so that  $\langle \mathcal{O} \rangle_t$  as defined above could also be written  $\langle \mathcal{O}(\mathcal{C}_t) \rangle$ .

The process associated to the current observable  $j_{\mathcal{O}}(\mathcal{C})$  at time  $t$  has a clear probabilistic meaning. Without explaining the details, let us just say that the decomposition

$$\mathcal{O}(\mathcal{C}_t) = \left( \mathcal{O}(\mathcal{C}_t) - \int_0^t ds j_{\mathcal{O}}(\mathcal{C}_s) \right) + \left( \int_0^t ds j_{\mathcal{O}}(\mathcal{C}_s) \right) \equiv M_t + N_t, \quad (3.14)$$

called the Doob-Meyer decomposition of the process  $\mathcal{O}(\mathcal{C}_t)$ , expresses  $\mathcal{O}(\mathcal{C}_t)$  as the sum of a martingale  $M_t$  and a predictable process  $N_t$  vanishing at  $t = 0$ . Such a decomposition is unique. Informally a martingale is a process whose expectation in the future knowing all the past up to

now is equal to its present value. In particular, the expectation  $\langle M_t \rangle$  is equal to 0. Taking this as a fact, one gets immediately (3.13). Informally again,  $N_t$  is predictable because its value at  $t + dt$ ,  $N_{t+dt}$ , is not sensitive to the randomness (i.e. to a possible jump occurring) between  $t$  and  $t + dt$ .

Even if  $\mathcal{O}(\mathcal{C}_t)$  depends only of the configuration at time  $t$ , this is not true anymore of  $M_t$  and  $N_t$  which in general depend on the history up to time  $t$ . Moreover, the decomposition does not behave trivially under nonlinear maps, so that  $N_t^2$  is not the predictable process that appears in the Doob-Meyer decomposition of  $[\mathcal{O}(\mathcal{C}_t)]^2$ .

### 3.2.3 Mean energy time derivative

As an example of (3.13), the time derivative of the mean energy  $\langle \mathcal{E} \rangle_t$  is equal to the mean value of the energy current  $j_{\mathcal{E}}(\mathcal{C})$ . In the present model, energy variations are only due to heat exchanges with the two thermostats according to the conservation rule

$$\mathcal{E}(\mathcal{C}') - \mathcal{E}(\mathcal{C}) = \delta q_1(\mathcal{C}' \leftarrow \mathcal{C}) + \delta q_2(\mathcal{C}' \leftarrow \mathcal{C}), \quad (3.15)$$

where  $\delta q_a(\mathcal{C}' \leftarrow \mathcal{C})$  is the heat received from thermal bath  $a$  for a jump of configuration as defined in (2.20). Therefore the energy current  $j_{\mathcal{E}}(\mathcal{C})$  can be split into two heat currents  $j_{\delta q_1}(\mathcal{C})$  and  $j_{\delta q_2}(\mathcal{C})$  received from the thermal baths 1 and 2 respectively. Then the evolution equation (3.13) applied to the energy observable  $\mathcal{E}$  can be rewritten as

$$\frac{d\langle \mathcal{E} \rangle_t}{dt} = \langle j_1 \rangle_t + \langle j_2 \rangle_t, \quad (3.16)$$

where  $j_a$  is a short notation for the instantaneous heat current received from the thermal bath  $a$  by the system when it leaves configuration  $\mathcal{C}$  :

$$j_a(\mathcal{C}) \equiv j_{\delta q_a}(\mathcal{C}) \equiv \sum_{\mathcal{C}'} (\mathcal{C}' | \mathbb{W} | \mathcal{C}) \delta q_a(\mathcal{C}' \leftarrow \mathcal{C}). \quad (3.17)$$

Note that, unless  $\beta_1 = \beta_2$ , there is no observable  $\mathcal{O}_a$  for which  $j_{\delta q_a}(\mathcal{C})$  would be equal to  $j_{\mathcal{O}_a}(\mathcal{C})$  with the definition of  $j_{\mathcal{O}_a}(\mathcal{C})$  given in (3.10). Moreover, under the microscopic reversibility condition (1.2), if  $\beta_1 = \beta_2 = \beta$  the modified detailed balance (2.16) becomes the canonical detailed balance (2.18). Then the current associated with any observable  $\mathcal{O}_a$  (more generally any exchange quantity) has a zero mean in the stationary equilibrium state with distribution  $P_{\text{can}}^\beta : \langle j_{\mathcal{O}_a} \rangle_{P_{\text{can}}^\beta} = 0$ .

### 3.2.4 Mean exchange entropy flow

The heat currents are associated to an exchange entropy variation of the system (see footnote 3 for the meaning) according to the relation (2.23). Similarly to the definition (3.17) of the microscopic heat current  $j_a(\mathcal{C})$  in terms of the heat amount  $\delta q_a(\mathcal{C}' \leftarrow \mathcal{C})$ , the microscopic exchange entropy current  $j_{\delta_{\text{exch}} S}(\mathcal{C})$  when the system goes out of the configuration  $\mathcal{C}$  is defined from the exchange entropy variation  $\delta_{\text{exch}} S(\mathcal{C}' \leftarrow \mathcal{C})$  as

$$j_{\delta_{\text{exch}} S}(\mathcal{C}) \equiv \sum_{\mathcal{C}'} (\mathcal{C}' | \mathbb{W} | \mathcal{C}) \delta_{\text{exch}} S(\mathcal{C}' \leftarrow \mathcal{C}). \quad (3.18)$$

The mean exchange entropy flow is defined as

$$\frac{d_{\text{exch}} S}{dt} \equiv \sum_{\mathcal{C}, \mathcal{C}'} (\mathcal{C}' | \mathbb{W} | \mathcal{C}) \delta_{\text{exch}} S(\mathcal{C}' \leftarrow \mathcal{C}) P(\mathcal{C}; t). \quad (3.19)$$

It can be expressed as the average of the current  $j_{\delta_{\text{exch}} S}(\mathcal{C})$ ,

$$\frac{d_{\text{exch}} S}{dt} = \langle j_{\delta_{\text{exch}} S} \rangle_t. \quad (3.20)$$

The current definitions (3.17) and (3.18) together with the relation (2.23) between the exchange entropy variation and heat transfers imply that

$$\frac{d_{\text{exch}}S}{dt} = \beta_1 \langle j_1 \rangle_t + \beta_2 \langle j_2 \rangle_t. \quad (3.21)$$

In the stationary state, the time derivative of the mean energy vanishes, namely  $d \langle \mathcal{E} \rangle_{\text{st}} / dt = 0$ , where  $\langle \dots \rangle_{\text{st}}$  denotes an average with the stationary distribution  $P_{\text{st}}$ . Henceforth, according to the evolution equation (3.16) of  $\langle \mathcal{E} \rangle_{\text{st}}$ ,

$$\langle j_1 \rangle_{\text{st}} + \langle j_2 \rangle_{\text{st}} = 0. \quad (3.22)$$

By inserting the current balance (3.22) into the expression (3.21) of the mean exchange entropy flow, we get

$$\left. \frac{d_{\text{exch}}S}{dt} \right|_{\text{st}} = -(\beta_1 - \beta_2) \langle j_2 \rangle_{\text{st}}. \quad (3.23)$$

### 3.3 Evolution of the Shannon-Gibbs entropy

#### 3.3.1 Definition of the entropy production rate

The dimensionless Shannon-Gibbs entropy (where the Boltzmann constant is set equal to 1) is defined from the configuration probability distribution  $P(\mathcal{C}; t)$  as

$$S^{SG}[P(t)] \equiv - \sum_{\mathcal{C}} P(\mathcal{C}; t) \ln P(\mathcal{C}; t) = - \langle \ln P(t) \rangle_t. \quad (3.24)$$

When the evolution is ruled by the master equation (3.1), the time derivative of  $S^{SG}(t)$  takes the form

$$\frac{dS^{SG}}{dt} = - \sum_{\mathcal{C}, \mathcal{C}'} (\mathcal{C}' | \mathbb{W} | \mathcal{C}) P(\mathcal{C}; t) \ln \frac{P(\mathcal{C}'; t)}{P(\mathcal{C}; t)} = - \langle j_{\ln P(t)} \rangle_t, \quad (3.25)$$

where we have used the definition (3.10) of an observable current  $j_{\mathcal{O}}(\mathcal{C}; t)$ , also valid in the case of an observable which depends explicitly on time.

As it is done for the phenomenological entropy introduced in the thermodynamic of irreversible processes [5, 55] the time derivative of  $S^{SG}(t)$  can be split into two contributions, an exchange (or external) part  $d_{\text{exch}}S/dt$  arising from exchanges with the external reservoirs, and an internal (or irreversible) part  $d_{\text{int}}S^{SG}/dt$  due to the internal irreversible processes in the system,

$$\frac{dS^{SG}}{dt} \equiv \frac{d_{\text{exch}}S}{dt} + \frac{d_{\text{int}}S^{SG}}{dt}. \quad (3.26)$$

By virtue of its definition (3.19),  $d_{\text{exch}}S/dt$  is expressed in terms of the exchange entropy variation  $\delta_{\text{exch}}S(\mathcal{C}' \leftarrow \mathcal{C})$ , associated with a jump of the system from a microscopic configuration  $\mathcal{C}$  to another one  $\mathcal{C}'$ . By virtue of the definition (3.26)  $d_{\text{int}}S^{SG}/dt$  is a functional of  $P(\mathcal{C}; t)$  determined from  $dS^{SG}/dt$  given in (3.25) and  $d_{\text{exch}}S/dt$  given in (3.19).

#### 3.3.2 Comparison with the thermodynamics of irreversible processes

An implicit postulate in the modern literature is that, when the system is out of equilibrium and evolves on time scales far smaller than the reservoirs, one can still define some universe entropy, which coincides with its statistical thermodynamic expression when the system and the reservoirs are at equilibrium, and where in the course of time the equilibrium Gibbs entropy of the system is replaced by its instantaneous Shannon-Gibbs entropy. In other words, the out-of-equilibrium universe entropy variation is the sum of the variation of the Shannon-Gibbs entropy of the system and the variation of the total thermodynamic entropy of the reservoirs, namely

$$\frac{dS_{\text{Univ}}}{dt} = \frac{dS^{SG}}{dt} + \frac{dS_{\text{res}}^{TH}}{dt}. \quad (3.27)$$

On the other hand, by definition, the exchange entropy flow  $d_{\text{exch}}S/dt$  received by the system from the reservoirs is the opposite of the time-derivative of the thermodynamic entropy of the reservoirs  $d_{\text{exch}}S/dt \equiv -dS_{\text{res}}^{TH}/dt$ . As a consequence, by virtue of definition (3.26),

$$\frac{dS_{\text{Univ}}}{dt} = \frac{d_{\text{int}}S^{SG}}{dt}. \quad (3.28)$$

The interpretation of the latter equality is that, since the reservoirs are at thermodynamic equilibrium, the variation rate of the universe entropy,  $dS_{\text{Univ}}/dt$ , is equal to the production rate of the Shannon-Gibbs entropy in the system, namely the internal part  $d_{\text{int}}S^{SG}/dt$  in the time-derivative  $dS^{SG}/dt$  of the system Shannon-Gibbs entropy. As shown below, if the modified detailed balance is obeyed, the universe entropy increases when the system is out-of-equilibrium, as the universe entropy increases between two equilibrium states, according to the second principle of thermodynamics.

In the stationary state the time derivative of the Shannon-Gibbs entropy (3.24) also vanishes,  $dS^{SG}[P_{\text{st}}]/dt = 0$ , and the decomposition (3.26) leads to

$$\left. \frac{d_{\text{int}}S^{SG}}{dt} \right|_{\text{st}} = - \left. \frac{d_{\text{exch}}S}{dt} \right|_{\text{st}}. \quad (3.29)$$

Then according to (3.23),  $d_{\text{int}}S^{SG}/dt|_{\text{st}}$  can be written as the entropy production rate introduced in the framework of irreversible thermodynamics when there is only one independent mean instantaneous current  $J$  (see for instance Ref.[56]), namely in the form

$$\left. \frac{d_{\text{int}}S^{SG}}{dt} \right|_{\text{st}} = \mathcal{F}J, \quad (3.30)$$

where  $\mathcal{F}$  is the so-called thermodynamic force associated with the heat current  $J$ . In the present case  $\langle j_2 \rangle_{\text{st}} = -\langle j_1 \rangle_{\text{st}}$  and in (3.30) we can make the identification

$$J = \langle j_2 \rangle_{\text{st}} \quad \text{and} \quad \mathcal{F} = \beta_1 - \beta_2. \quad (3.31)$$

(Another identification might have been  $J = \langle j_1 \rangle_{\text{st}}$  and  $\mathcal{F} = \beta_2 - \beta_1$ .) If  $T_2 > T_1$ ,  $\mathcal{F}$  is positive and the positivity of  $d_{\text{int}}S_{\text{st}}^{SG}/dt$  (settled in next subsection when the modified detailed balance is satisfied) ensures that the mean current  $\langle j_2 \rangle_{\text{st}}$  received from the heat bath 2 is also positive.

### 3.3.3 Entropy production rate under MDB

When the transition rates obey the modified detailed balance (2.24), the microscopic exchange entropy current defined in (3.18) can be rewritten as

$$j_{\delta_{\text{exch}}S}(\mathcal{C}) \stackrel{\text{MDB}}{=} - \sum_{\mathcal{C}'} (\mathcal{C}'|\mathbb{W}|\mathcal{C}) \ln \frac{(\mathcal{C}'|\mathbb{W}|\mathcal{C})}{(\mathcal{C}|\mathbb{W}|\mathcal{C}')}. \quad (3.32)$$

Then the exchange entropy flow defined in (3.19) becomes equal to

$$\frac{d_{\text{exch}}S}{dt} \stackrel{\text{MDB}}{=} - \sum_{\mathcal{C}, \mathcal{C}'} (\mathcal{C}'|\mathbb{W}|\mathcal{C}) P(\mathcal{C}; t) \ln \frac{(\mathcal{C}'|\mathbb{W}|\mathcal{C})}{(\mathcal{C}|\mathbb{W}|\mathcal{C}')} \equiv \sigma_{\text{exch}}[P(t)], \quad (3.33)$$

while the entropy production rate given by (3.25) and (3.26) becomes equal to

$$\frac{d_{\text{int}}S^{SG}}{dt} \stackrel{\text{MDB}}{=} \sum_{\mathcal{C}, \mathcal{C}'} (\mathcal{C}'|\mathbb{W}|\mathcal{C}) P(\mathcal{C}; t) \ln \frac{(\mathcal{C}'|\mathbb{W}|\mathcal{C}) P(\mathcal{C}; t)}{(\mathcal{C}|\mathbb{W}|\mathcal{C}') P(\mathcal{C}'; t)} \equiv \sigma_{\text{int}}[P(t)] \quad (3.34)$$

The latter expression takes the form of a relative entropy so that  $d_{\text{int}}S^{SG}/dt$  is positive under MDB.

We notice that, in the generic case where the modified detailed balance does not necessarily holds, Lebowitz and Spohn [1] have introduced the splitting of  $dS^{SG}/dt$  into  $\sigma_{\text{exch}} + \sigma_{\text{int}}$ , where  $\sigma_{\text{exch}}[P(t)]$  is defined in (3.33) and  $\sigma_{\text{int}}[P(t)]$  is defined as the symmetrized expression of the definition in (3.34),

$$\sigma_{\text{int}}[P(t)] = \frac{1}{2} \sum_{\mathcal{C}, \mathcal{C}'} [(\mathcal{C}'|\mathbb{W}|\mathcal{C})P(\mathcal{C};t) - (\mathcal{C}|\mathbb{W}|\mathcal{C}')P(\mathcal{C}';t)] \ln \frac{(\mathcal{C}'|\mathbb{W}|\mathcal{C})P(\mathcal{C};t)}{(\mathcal{C}|\mathbb{W}|\mathcal{C}')P(\mathcal{C}';t)}, \quad (3.35)$$

which is obviously positive. Another rewriting for the definition of  $\sigma_{\text{int}}[P(t)]$  given in (3.34) is

$$\sigma_{\text{int}}[P(t)] = \langle j_{\delta_{\text{int}}S}(t) \rangle_t \quad \text{with} \quad \delta_{\text{int}}S(\mathcal{C}' \leftarrow \mathcal{C}) \equiv \ln \frac{(\mathcal{C}'|\mathbb{W}|\mathcal{C})P(\mathcal{C};t)}{(\mathcal{C}|\mathbb{W}|\mathcal{C}')P(\mathcal{C}';t)}. \quad (3.36)$$

Then the positivity of  $\sigma_{\text{int}}[P(t)]$  may be viewed as arising from the property : for  $x > 0$ ,  $-\ln x \geq 1 - x$ . We notice that  $j_{\delta_{\text{int}}S}(\mathcal{C};t)$  is not the current introduced in [1]. However, the current  $j_{\delta_{\text{exch}}S}(\mathcal{C})$  is introduced through its expression (3.32) there and it is used again, under another denomination, in [57].

### 3.3.4 Decomposition of the entropy production rate into two positive contributions

The microscopic variation  $\delta_{\text{int}}S(\mathcal{C}' \leftarrow \mathcal{C})$  defined in (3.36) can be decomposed into two contributions as

$$\delta_{\text{int}}S(\mathcal{C}' \leftarrow \mathcal{C}) = - \left[ \ln \frac{P(\mathcal{C}';t)}{P_{\text{st}}(\mathcal{C}')} - \ln \frac{P(\mathcal{C};t)}{P_{\text{st}}(\mathcal{C})} \right] + \delta A_{\text{st}}(\mathcal{C}' \leftarrow \mathcal{C}), \quad (3.37)$$

where

$$\delta A_{\text{st}}(\mathcal{C}' \leftarrow \mathcal{C}) \equiv \ln \frac{(\mathcal{C}'|\mathbb{W}|\mathcal{C})P_{\text{st}}(\mathcal{C})}{(\mathcal{C}|\mathbb{W}|\mathcal{C}')P_{\text{st}}(\mathcal{C}')}.$$

By analogy with chemical reaction kinetics (see for instance Ref.[52] and subsection 3.3.5 below),  $\delta A_{\text{st}}(\mathcal{C}' \leftarrow \mathcal{C})$  may be viewed as the affinity of the elementary reversible reaction (or phase change)  $\mathcal{C} \rightleftharpoons \mathcal{C}'$  in the stationary state resulting from all possible reversible pair reactions between all configurations. We notice that  $\delta_{\text{int}}S(\mathcal{C}' \leftarrow \mathcal{C})$  depends explicitly on time, contrarily to  $\delta A_{\text{st}}(\mathcal{C}' \leftarrow \mathcal{C})$ . According to the definitions (3.10) and (3.11) of the currents associated respectively to an observable  $\mathcal{O}$  or to some exchange quantity  $\delta K(\mathcal{C}' \leftarrow \mathcal{C})$ , the decomposition (3.37) allows to rewrite the expression (3.36) for  $\sigma_{\text{int}}[P(t)]$  as

$$\sigma_{\text{int}}[P(t)] = -\langle j_{\ln[P(t)/P_{\text{st}}]} \rangle_t + \langle j_{\delta A_{\text{st}}} \rangle_t. \quad (3.39)$$

We stress that, by virtue of the already used inequality  $-\ln x \geq 1 - x$  for  $x > 0$  and according to the stationary condition  $\sum_{\mathcal{C}' \neq \mathcal{C}} (\mathcal{C}|\mathbb{W}|\mathcal{C}')P_{\text{st}}(\mathcal{C}') - \sum_{\mathcal{C}' \neq \mathcal{C}} (\mathcal{C}'|\mathbb{W}|\mathcal{C})P_{\text{st}}(\mathcal{C}) = 0$  (see (3.1)), the microscopic current  $j_{\delta A_{\text{st}}}(\mathcal{C})$  which we have introduced is positive,

$$j_{\delta A_{\text{st}}}(\mathcal{C}) \geq 0 \quad \text{for any } \mathcal{C}. \quad (3.40)$$

The latter positivity may be viewed as a consequence of the fact that  $j_{\delta A_{\text{st}}}(\mathcal{C})$  is the relative entropy (for a given  $\mathcal{C}$ ) of the rate  $(\mathcal{C}'|\mathbb{W}|\mathcal{C})$  with respect to the rate  $(\mathcal{C}'|\mathbb{W}^\dagger|\mathcal{C}) \equiv [P_{\text{st}}(\mathcal{C})]^{-1} (\mathcal{C}|\mathbb{W}|\mathcal{C}')P_{\text{st}}(\mathcal{C}')$ .

According to the master equation (3.1), the first contribution in the r.h.s. of (3.39) is equal to the opposite of the time derivative of the relative entropy of the probability distribution  $P(t)$  with respect to the stationary solution defined in (3.5). Therefore the expression (3.36) of the entropy production rate  $\sigma_{\text{int}}[P(t)]$  can be split into two contributions, both of which are positive,

$$\sigma_{\text{int}}[P(t)] = -\frac{dS_{\text{rel}}[P(t)|P_{\text{st}}]}{dt} + \langle j_{\delta A_{\text{st}}} \rangle_t. \quad (3.41)$$

The positivity of the first term has played a role in the discussion of the uniqueness of the stationary state (see Eq.(3.9)) and the positivity of the second term arises from the positivity of the current pointed out in (3.40).

For the sake of completeness, we notice that, when the transition rates depend on time the master equation (3.1) remains unchanged, so that the decomposition of the time-derivative of the Shannon-Gibbs entropy into  $\sigma_{\text{exch}} + \sigma_{\text{int}}$  as well as the decomposition (3.39) of  $\sigma_{\text{int}}$  remain unchanged, except that  $P_{\text{st}}$  is replaced by a function of time, namely the probability distribution which is the zero eigenvector of the Markov matrix at a given time. Then the splitting (3.39) of  $\sigma_{\text{int}}$  is quite analogous to that introduced in [39] for an evolution where both the bath temperatures and the energy levels may change with time. The so-called non-adiabatic term corresponds to the opposite of the time-derivative of the relative entropy  $S_{\text{rel}}[P(t)|P_{\text{st}}]$  and the so-called adiabatic term corresponds to the average of the stationary affinity current.

### 3.3.5 Comparison with NESS characterization from graph theory

The r.h.s. of the master equation (3.1) can be rewritten as a sum of probability currents between configurations,

$$\frac{dP(\mathcal{C}; t)}{dt} = \sum_{\mathcal{C}, \mathcal{C}'} J_{[\mathcal{P}]}(\mathcal{C}', \mathcal{C}; t), \quad (3.42)$$

where

$$J_{[\mathcal{P}]}(\mathcal{C}', \mathcal{C}; t) \equiv (\mathcal{C}'|\mathbb{W}|\mathcal{C})P(\mathcal{C}; t) - (\mathcal{C}|\mathbb{W}|\mathcal{C}')P(\mathcal{C}'; t). \quad (3.43)$$

The probability current  $J_{[\mathcal{P}]}(\mathcal{C}', \mathcal{C}; t)$  has the form of the chemical reaction rate associated with the reversible reaction (or phase change)  $\mathcal{C} \rightleftharpoons \mathcal{C}'$  with the species concentrations replaced by the configuration probabilities  $P(\mathcal{C})$  and  $P(\mathcal{C}')$  and the reaction rate constant for  $\mathcal{C} \rightarrow \mathcal{C}'$  replaced by the transition rate  $(\mathcal{C}'|\mathbb{W}|\mathcal{C})$ . The entropy production rate  $\sigma_{\text{int}}[P(t)]$ , introduced in Ref.[1] in the generic case (namely when the MDB is not necessarily satisfied) by the definition (3.35), can be rewritten (with notations similar to those of Ref.[52]) as

$$\sigma_{\text{int}}[P(t)] = \frac{1}{2} \sum_{\mathcal{C}, \mathcal{C}'} \delta A_{[\mathcal{P}]}(\mathcal{C}', \mathcal{C}; t) J_{[\mathcal{P}]}(\mathcal{C}', \mathcal{C}; t), \quad (3.44)$$

where the dimensionless affinity  $\delta A_{[\mathcal{P}]}(\mathcal{C}', \mathcal{C}; t)$  of an oriented pair (already introduced in the case of the stationary distribution  $P_{\text{st}}$  in (3.38) with a slightly different notation) is defined as

$$\delta A_{[\mathcal{P}]}(\mathcal{C}', \mathcal{C}; t) \equiv \ln \frac{(\mathcal{C}'|\mathbb{W}|\mathcal{C})P(\mathcal{C}; t)}{(\mathcal{C}|\mathbb{W}|\mathcal{C}')P(\mathcal{C}'; t)}. \quad (3.45)$$

Indeed, in irreversible thermodynamics of chemical reactions at a temperature  $T$  fixed by a thermostat (see for instance [58]), if  $P(\mathcal{C})$  denotes the ratio between the concentration of species  $\mathcal{C}$  and the sum of all species concentrations, the affinity of the reaction  $\mathcal{C} \rightleftharpoons \mathcal{C}'$  is equal to  $k_B T$  times an expression similar to  $\delta A_{[\mathcal{P}]}(\mathcal{C}', \mathcal{C}; t)$ , with the concentration of species  $\mathcal{C}$  ( $\mathcal{C}'$ ) in place of  $P(\mathcal{C}; t)$  ( $P(\mathcal{C}'; t)$ ) and the reaction rate constant for  $\mathcal{C} \rightarrow \mathcal{C}'$  in place of the transition rate  $(\mathcal{C}'|\mathbb{W}|\mathcal{C})$ .

The previous rewritings are convenient to handle the master equation in the framework of network theory where the master equation is represented by a graph  $G$  as follows. Each vertex of the graph corresponds to a given configuration  $\mathcal{C}$  and there exists an edge between two vertices if at least one of the transition rates  $(\mathcal{C}'|\mathbb{W}|\mathcal{C})$  or  $(\mathcal{C}|\mathbb{W}|\mathcal{C}')$  does not vanish. Moreover an arbitrary orientation is chosen for every edge in the graph  $G$  so that the transition rate  $(\mathcal{C}'|\mathbb{W}|\mathcal{C})$  (or  $(\mathcal{C}|\mathbb{W}|\mathcal{C}')$ ) can be shortly referred to as the transition rate along the edge in either the positive or the negative sense.

From a connected graph  $G$  one can define several possible fundamental sets of  $N_c$  circuits (or closed paths) on the graph. The number  $N_c$  of circuits is only determined by the edge number  $N_e$  and the vertex number  $N_v$  through the relation  $N_c = N_e - N_v + 1$ . (The final results are independent of the specific fundamental set used in intermediate algebraic calculations.) A fundamental set is built from one the various possible maximal (or spanning) trees which are defined by removing  $N_c$  edges of  $G$ . For a given maximal tree  $T(G)$  the corresponding removed edges are called chords and indexed by  $\alpha = 1, \dots, N_c$ . For each  $\alpha$  the circuit  $\mathcal{C}_\alpha$  is obtained by first considering the graph made by adding the chord  $\alpha$  to  $T(G)$  and then removing all edges which are not part of the circuit



closed by the insertion of the chord  $\alpha$ . A cycle  $\vec{C}_\alpha$  is associated with every circuit  $C_\alpha$  by choosing an arbitrary orientation to go around the circuit, so “cycle” is a synonymous for “oriented circuit”.

The affinity  $A(\vec{C}_\alpha)$  associated to a cycle  $\vec{C}_\alpha$  is defined as an algebraic sum of all edge affinities which is calculated as follows: each edge affinity, whose expression is given by (3.45) for a positive orientation from  $\mathcal{C}$  to  $\mathcal{C}'$  in the graph  $G$ , is multiplied by the sign of the relative orientation of the edge in the cycle  $\vec{C}_\alpha$  and in the graph  $G$ . Because of the cyclic structure of  $\vec{C}_\alpha$  the affinity  $A(\vec{C}_\alpha)$  does not depend explicitly on the configuration probability distribution  $P$  and it is determined only from the transition rates  $(\mathcal{C}'|\mathbb{W}|\mathcal{C})$ . Indeed, if  $N(\vec{C}_\alpha)$  is the number of configurations involved in the cycle  $\vec{C}_\alpha$ , and if the configurations are labeled with indices increasing with one unit when one goes from one configuration to the next one in the sense chosen for the orientation of the cycle  $\vec{C}_\alpha$ , then

$$A(\vec{C}_\alpha) = \ln \prod_{i=1}^{N(\vec{C}_\alpha)} \frac{(\mathcal{C}_{i+1}|\mathbb{W}|\mathcal{C}_i)}{(\mathcal{C}_i|\mathbb{W}|\mathcal{C}_{i+1})}, \quad (3.46)$$

with the notational convention  $\mathcal{C}_{N(\vec{C}_\alpha)+1} \equiv \mathcal{C}_1$ . The probability current  $J_{[P]}(\vec{C}_\alpha; t)$  associated to the cycle is defined as the probability current in the chord  $\alpha$  in the sense of the cycle orientation : it is given by definition (3.43) where  $\mathcal{C}$  and  $\mathcal{C}'$  are respectively the initial and final configurations in the sense of the cycle orientation.

As shown in Ref.[52], the stationary state has the following properties. First, the vanishing of all cycle affinities  $A(\vec{C}_\alpha)$  is equivalent to the vanishing of all cycle probability currents  $J_{[P_{\text{st}}]}(\vec{C}_\alpha)$  in the stationary state,

$$\forall \alpha \quad A(\vec{C}_\alpha) = 0 \quad \Leftrightarrow \quad \forall \alpha \quad J_{[P_{\text{st}}]}(\vec{C}_\alpha) = 0. \quad (3.47)$$

Second, the vanishing of all stationary cycle probability currents  $J_{[P_{\text{st}}]}(\vec{C}_\alpha)$  is equivalent to the fact that the stationary state obeys the detailed balance condition, namely

$$\forall \alpha \quad J_{[P_{\text{st}}]}(\vec{C}_\alpha) = 0 \quad \Leftrightarrow \quad \forall (\mathcal{C}, \mathcal{C}') \quad \frac{(\mathcal{C}'|\mathbb{W}|\mathcal{C})P_{\text{st}}(\mathcal{C})}{(\mathcal{C}|\mathbb{W}|\mathcal{C}')P_{\text{st}}(\mathcal{C}')} = 1. \quad (3.48)$$

Schnakenberg specifies that it is in fact a “complete detailed balance” condition in the sense that if there exist several kinds of independent transitions between two configurations  $\mathcal{C}$  and  $\mathcal{C}'$ , the detailed balance must be satisfied by every kind of transition. Since equilibrium is characterized on the mesoscopic level by the detailed balance (see the derivation of (2.18)), the properties (3.47) and (3.48) entails that the equilibrium is characterized by either the vanishing of all stationary cycle currents or the vanishing of all cycle affinities. Moreover, the entropy production rate (3.44) in the stationary state reads

$$\sigma_{\text{int}}[P_{\text{st}}] = \sum_{\alpha=1}^{N_c} A(\vec{C}_\alpha) J_{[P_{\text{st}}]}(\vec{C}_\alpha), \quad (3.49)$$

where  $A(\vec{C}_\alpha)$  is given directly in terms of the transition rates by the formula (3.46). By virtue of the definitions (3.33) and (3.34),  $\sigma_{\text{exch}}[P(t)] + \sigma_{\text{int}}[P(t)] = dS^{SG}/dt$  and in the stationary state  $\sigma_{\text{int}}[P_{\text{st}}] = -\sigma_{\text{exch}}[P_{\text{st}}]$ .

When the MDB is satisfied, according to (3.33),  $\sigma_{\text{exch}}[P_{\text{st}}] \stackrel{\text{MDB}}{=} d_{\text{exch}}S/dt|_{\text{st}}$  so that  $\sigma_{\text{int}}[P_{\text{st}}] \stackrel{\text{MDB}}{=} -d_{\text{exch}}S/dt|_{\text{st}}$  and (3.49) becomes

$$\left. \frac{d_{\text{exch}}S}{dt} \right|_{\text{st}} \stackrel{\text{MDB}}{=} - \sum_{\alpha=1}^{N_c} A(\vec{C}_\alpha) J_{[P_{\text{st}}]}(\vec{C}_\alpha). \quad (3.50)$$

Meanwhile, by virtue of (1.3) the affinity of a cycle  $A(\vec{C}_\alpha)$  given by (3.46) becomes equal to

$$A(\vec{C}_\alpha) \stackrel{\text{MDB}}{=} - \sum_{i=1}^{N(\vec{C}_\alpha)} \delta_{\text{exch}}S(\mathcal{C}_{i+1} \leftarrow \mathcal{C}_i), \quad (3.51)$$

where the latter expression depends on the thermodynamic parameters of the reservoirs and on the quanta of microscopic quantities which the system exchanges with the reservoirs. In the generic case there may be several cycles corresponding to the same current between two reservoirs and there is no straightforward correspondence between the expression of  $-d_{\text{exch}}S/dt|_{\text{st}}$  given by (3.50) in terms of the  $A(\vec{C}_\alpha)$ 's and  $J_{[P_{\text{st}}]}(\vec{C}_\alpha)$ 's and the expression of  $-d_{\text{exch}}S/dt|_{\text{st}}$  given by the entropy production rate for the phenomenological entropy  $S$  in thermodynamics of irreversible processes  $-d_{\text{exch}}S/dt|_{\text{st}} = d_{\text{int}}S/dt|_{\text{st}} = \sum_\gamma \mathcal{F}_\gamma J_\gamma^*$ , where the  $J_\gamma^*$ 's are independent macroscopic currents.

However in the case of the simple solvable model considered in paper II, where the thermal contact between two thermostats is settled by a two-spin system, only one cycle is involved and one can make the correspondence between, on the one hand, the affinity  $A(\vec{C})$  of the cycle and the stationary current  $J_{[P_{\text{st}}]}(\vec{C})$  which goes through it, and, on the other hand, the thermodynamic force  $\mathcal{F}$  and the mean heat current  $J$ . Indeed, in this model, the graph associated to the master equation is itself a cycle, which can be orientated to read

$$\begin{array}{ccc} (+, +) & \rightarrow & (-, +) \\ \uparrow & & \downarrow \\ (+, -) & \leftarrow & (-, -) \end{array} . \quad (3.52)$$

The cycle may be rewritten as  $\mathcal{C}_1 \rightarrow \mathcal{C}_2 \rightarrow \mathcal{C}_3 \rightarrow \mathcal{C}_4 \rightarrow \mathcal{C}_1$  where the configurations are labeled in the positive sense of the cycle orientation. Then, since the model obeys the MDB, the affinity of the cycle is given by (3.51), and by virtue of the definitions (2.20) and (2.23) it reads  $A(\vec{C}) = -\beta_1 q_1(\vec{C}) - \beta_2 q_2(\vec{C})$ , where  $q_a(\vec{C})$  is the heat received from the thermostat  $a$  when the system configurations perform the cycle once in the positive sense. Moreover, according to the energy conservation law,  $q_1(\vec{C}) + q_2(\vec{C})$  is equal to the energy difference between the final and initial states when the cycle is performed once : this difference vanishes for a cycle so that

$$A(\vec{C}) = (\beta_1 - \beta_2) q_2(\vec{C}). \quad (3.53)$$

On the other hand, since the graph is exactly a cycle, the current along an edge defined in (3.43) has the same value for all edges in the stationary state (because, by virtue of (3.42), the stationary condition  $dP(\mathcal{C}; t)/dt|_{\text{st}} = 0$  is equivalent to Kirchoff's current law at every vertex of the graph). As a consequence, the stationary current associated to the cycle reads

$$J_{[P_{\text{st}}]}(\vec{C}) = (\mathcal{C}_{i+1}|\mathbb{W}|\mathcal{C}_i)P_{\text{st}}(\mathcal{C}_i) - (\mathcal{C}_i|\mathbb{W}|\mathcal{C}_{i+1})P_{\text{st}}(\mathcal{C}_{i+1}), \quad (3.54)$$

where  $i$  is any label in  $\{1, 2, 3, 4\}$ . The mean exchange entropy flow given by (3.50) reads  $d_{\text{exch}}S/dt|_{\text{st}} = -(\beta_1 - \beta_2) q_2(\vec{C}) J_{[P_{\text{st}}]}(\vec{C})$ . Comparison with the expression  $d_{\text{exch}}S/dt = -\mathcal{F}J$  of irreversible processes thermodynamics (see (3.29)-(3.30)-(3.31)), leads to the following identification of the macroscopic heat current  $J = \langle j_2 \rangle_{\text{st}}$  received from the heat bath 2 (and given to the heat bath 1),

$$J = q_2(\vec{C}) J_{[P_{\text{st}}]}(\vec{C}), \quad (3.55)$$

while the thermodynamic force  $\mathcal{F} = \beta_1 - \beta_2$  is to be identified with

$$\mathcal{F} = \frac{A(\vec{C})}{q_2(\vec{C})}. \quad (3.56)$$

The probabilistic interpretation of the affinity  $A(\vec{C})$  is given in paper II [45].

## 4 Exchange entropy variation and symmetries at finite time under MDB

### 4.1 Exchange entropy variation for a history

For a history  $\mathcal{H}ist$  where the system starts in configuration  $\mathcal{C}_0$  at time  $t_0 = 0$  and ends in configuration  $\mathcal{C}_f$  at time  $t$  after going through successive configurations  $\mathcal{C}_0, \mathcal{C}_1, \dots, \mathcal{C}_N = \mathcal{C}_f$ , the exchange

entropy variation  $\Delta_{\text{exch}}S[\mathcal{H}ist]$  corresponding to the history is defined from the heat amounts  $\mathcal{Q}_1[\mathcal{H}ist]$  and  $\mathcal{Q}_2[\mathcal{H}ist]$  received from the thermal baths as

$$\Delta_{\text{exch}}S[\mathcal{H}ist] \equiv \beta_1 \mathcal{Q}_1[\mathcal{H}ist] + \beta_2 \mathcal{Q}_2[\mathcal{H}ist] \quad \text{with} \quad \mathcal{Q}_a[\mathcal{H}ist] \equiv \sum_{i=0}^{N-1} \delta q_a(\mathcal{C}_{i+1} \leftarrow \mathcal{C}_i). \quad (4.1)$$

The expectation value of  $\Delta_{\text{exch}}S[\mathcal{H}ist]$  with respect to the measure over all possible histories starting from configurations distributed according to some initial probability distribution (see appendix D) is equal to the time integral of the mean exchange entropy current calculated with the instantaneous configuration probability distribution,

$$\langle \Delta_{\text{exch}}S \rangle = \int_0^t dt' \langle j_{\delta_{\text{exch}}}S \rangle_{t'} = \int_0^t dt' \frac{d_{\text{exch}}S}{dt'}. \quad (4.2)$$

The second equality arises from (3.20).

## 4.2 MDB and symmetry between time-reversed histories

Let  $\mathbb{T}$  be the time reversal operator for histories. If  $\mathcal{H}ist$  is a history that starts at time  $t_0 = 0$  in  $\mathcal{C}_0$  and ends at time  $t$  in  $\mathcal{C}_f$  after  $N$  jumps from  $\mathcal{C}_{i-1}$  to  $\mathcal{C}_i$  at time  $T_i$ ,  $\mathbb{T}\mathcal{H}ist$  is a history that starts at time  $t_0 = 0$  in  $\mathcal{C}_f$  and ends at  $\mathcal{C}_0$  at time  $t$  after  $N$  jumps from  $\mathcal{C}'_{i-1}$  to  $\mathcal{C}'_i$  at time  $T'_i$  with  $\mathcal{C}'_i = \mathcal{C}_{N-i}$  and  $T'_i = t - T_{N-i+1}$ , namely

$$\begin{aligned} \mathcal{H}ist : \quad & \mathcal{C}_0 \text{ at } t_0 = 0 & \mathcal{C}_0 \xrightarrow{T_1} \mathcal{C}_1 \cdots \mathcal{C}_{N-1} \xrightarrow{T_N} \mathcal{C}_f \\ \mathbb{T}\mathcal{H}ist : \quad & \mathcal{C}_f \text{ at } t_0 = 0 & \mathcal{C}_f \xrightarrow{T'_1} \mathcal{C}_{N-1} \cdots \mathcal{C}_1 \xrightarrow{T'_N} \mathcal{C}_0. \end{aligned} \quad (4.3)$$

From the definition of the measure  $dP_{\mathcal{C}_0, \mathcal{C}_f}$  over histories starting in configuration  $\mathcal{C}_0$  and ending in configuration  $\mathcal{C}_f$  (see Appendix D),

$$\frac{dP_{\mathcal{C}_f, \mathcal{C}_0}[\mathcal{H}ist]}{dP_{\mathcal{C}_0, \mathcal{C}_f}[\mathbb{T}\mathcal{H}ist]} = \prod_{i=0}^{N-1} \frac{(\mathcal{C}_{i+1}|\mathbb{W}|\mathcal{C}_i)}{(\mathcal{C}_i|\mathbb{W}|\mathcal{C}_{i+1})}. \quad (4.4)$$

When the transition rates obey the modified detailed balance (2.24) written in terms of  $\delta_{\text{exch}}S(\mathcal{C}' \leftarrow \mathcal{C})$ , the exchange entropy variation for the history, defined in (4.1), can be rewritten as

$$\Delta_{\text{exch}}S[\mathcal{H}ist] \stackrel{\text{MDB}}{=} -\ln \prod_{i=0}^{N-1} \frac{(\mathcal{C}_{i+1}|\mathbb{W}|\mathcal{C}_i)}{(\mathcal{C}_i|\mathbb{W}|\mathcal{C}_{i+1})}, \quad (4.5)$$

and equation (4.4) can be rewritten as

$$\frac{dP_{\mathcal{C}_f, \mathcal{C}_0}[\mathcal{H}ist]}{dP_{\mathcal{C}_0, \mathcal{C}_f}[\mathbb{T}\mathcal{H}ist]} \stackrel{\text{MDB}}{=} e^{-\Delta_{\text{exch}}S[\mathcal{H}ist]}. \quad (4.6)$$

We stress that, according to (4.5), when the MDB is satisfied the expression of the exchange entropy variation for a history defined in (4.1) coincides with the opposite of the “action functional” introduced by Lebowitz and Spohn in Ref.[1] in the generic case where the MDB does not necessarily hold.

## 4.3 Symmetry between time-reversed evolutions with fixed heat amounts

The probability  $P(\mathcal{C}_f | \mathcal{Q}_1, \mathcal{Q}_2, t | \mathcal{C}_0)$  that the system has evolved from the configuration  $\mathcal{C}_0$  at  $t_0 = 0$  to the configuration  $\mathcal{C}_f$  at  $t$  while receiving the heat amounts  $\mathcal{Q}_1$  and  $\mathcal{Q}_2$  from the thermostats 1 and 2 reads

$$P(\mathcal{C}_f | \mathcal{Q}_1, \mathcal{Q}_2, t | \mathcal{C}_0) \equiv \int dP_{\mathcal{C}_f, \mathcal{C}_0}[\mathcal{H}ist] \delta(\mathcal{Q}_1[\mathcal{H}ist] - \mathcal{Q}_1) \delta(\mathcal{Q}_2[\mathcal{H}ist] - \mathcal{Q}_2), \quad (4.7)$$

where  $\int dP_{\mathcal{C}_f, \mathcal{C}_0}$  denotes the “summation” over the histories from  $\mathcal{C}_0$  to  $\mathcal{C}_f$ . The time-reversal symmetry property (4.6) for the history measure  $dP_{\mathcal{C}_f, \mathcal{C}_0} [\mathcal{H}ist]$  implies the following relation between probabilities of forward and backward evolutions where initial and final configurations are exchanged (and the heat amounts are changed into their opposite values),

$$\frac{P(\mathcal{C}_f | \mathcal{Q}_1, \mathcal{Q}_2; t | \mathcal{C}_0)}{P(\mathcal{C}_0 | -\mathcal{Q}_1, -\mathcal{Q}_2; t | \mathcal{C}_f)} = e^{-\Delta_{\text{exch}} S(\mathcal{Q}_1, \mathcal{Q}_2)}, \quad (4.8)$$

with the definition

$$\Delta_{\text{exch}} S(\mathcal{Q}_1, \mathcal{Q}_2) \equiv \beta_1 \mathcal{Q}_1 + \beta_2 \mathcal{Q}_2. \quad (4.9)$$

An analogous relation for  $\Delta_{\text{exch}} S$  in place of  $(\mathcal{Q}_1, \mathcal{Q}_2)$  is derived in [59] in the case where the microscopic dynamics of the heat baths are assumed to be Hamiltonian.

#### 4.4 Symmetries in protocols starting from an equilibrium state

We consider a protocol where the system is prepared in an equilibrium state at the inverse temperature  $\beta_0$  and then put in thermal contact with the two thermostats at the inverse temperatures  $\beta_1$  and  $\beta_2$  respectively at time  $t_0 = 0$ . Such a protocol is a transposed version of the Jarzynski procedure [30] where the system has been prepared in an equilibrium state at the initial time and then is submitted to a variation of some control parameter which drives the system out of equilibrium by injecting some work. In the present protocol, work is replaced by the heat flux that goes through the system from one thermal bath to the other, and the system evolution is a relaxation from an equilibrium state to a stationary non-equilibrium state.

The initial equilibrium distribution at the inverse temperature  $\beta_0$  is the canonical distribution (2.17).  $Z(\beta_0)$  cancels in the ratio  $P_{\text{can}}^{\beta_0}(\mathcal{C}_0)/P_{\text{can}}^{\beta_0}(\mathcal{C}_f)$  and

$$\ln \frac{P_{\text{can}}^{\beta_0}(\mathcal{C}_0)}{P_{\text{can}}^{\beta_0}(\mathcal{C}_f)} = \beta_0 [\mathcal{E}(\mathcal{C}_f) - \mathcal{E}(\mathcal{C}_0)] = \beta_0 (\mathcal{Q}_1 + \mathcal{Q}_2), \quad (4.10)$$

where the last equality is enforced by energy conservation. Then the time-reversal symmetry (4.8) and the specific form (4.10) for  $\ln[P_{\text{can}}^{\beta_0}(\mathcal{C}_0)/P_{\text{can}}^{\beta_0}(\mathcal{C}_f)]$  imply that

$$\frac{P(\mathcal{C}_f | \mathcal{Q}_1, \mathcal{Q}_2; t | \mathcal{C}_0) P_{\text{can}}^{\beta_0}(\mathcal{C}_0)}{P(\mathcal{C}_0 | -\mathcal{Q}_1, -\mathcal{Q}_2; t | \mathcal{C}_f) P_{\text{can}}^{\beta_0}(\mathcal{C}_f)} = e^{-\Delta_{\text{exch}}^{\text{excs}, \beta_0} S(\mathcal{Q}_1, \mathcal{Q}_2)}, \quad (4.11)$$

where  $\Delta_{\text{exch}}^{\text{excs}, \beta_0} S(\mathcal{Q}_1, \mathcal{Q}_2)$  is an excess exchange entropy variation defined as

$$\Delta_{\text{exch}}^{\text{excs}, \beta_0} S(\mathcal{Q}_1, \mathcal{Q}_2) = \Delta_{\text{exch}} S(\mathcal{Q}_1, \mathcal{Q}_2) - \beta_0 (\mathcal{Q}_1 + \mathcal{Q}_2) = (\beta_1 - \beta_0) \mathcal{Q}_1 + (\beta_2 - \beta_0) \mathcal{Q}_2. \quad (4.12)$$

$\Delta_{\text{exch}}^{\text{excs}, \beta_0} S(\mathcal{Q}_1, \mathcal{Q}_2)$  does not depend explicitly on the final and initial configurations and is only a function of the heat amounts received from the thermal baths,

As a consequence, the measurable joint distribution  $P_{P_{\text{can}}^{\beta_0}}(\mathcal{Q}_1, \mathcal{Q}_2; t)$  for the heat amounts  $\mathcal{Q}_1$  and  $\mathcal{Q}_2$  received between  $t_0 = 0$  and  $t$  when the initial configuration of the system is distributed according to  $P_{\text{can}}^{\beta_0}$ , which is defined as  $P_{P_{\text{can}}^{\beta_0}}(\mathcal{Q}_1, \mathcal{Q}_2; t) = \sum_{\mathcal{C}_0, \mathcal{C}_f} P(\mathcal{C}_f | \mathcal{Q}_1, \mathcal{Q}_2, t | \mathcal{C}_0) P_{\text{can}}^{\beta_0}(\mathcal{C}_0)$ , satisfies the identity

$$\frac{P_{P_{\text{can}}^{\beta_0}}(\mathcal{Q}_1, \mathcal{Q}_2; t)}{P_{P_{\text{can}}^{\beta_0}}(-\mathcal{Q}_1, -\mathcal{Q}_2; t)} = e^{-\Delta_{\text{exch}}^{\text{excs}, \beta_0} S(\mathcal{Q}_1, \mathcal{Q}_2)}. \quad (4.13)$$

Subsequently the measurable quantity  $\Delta_{\text{exch}}^{\text{excs}, \beta_0} S(\mathcal{Q}_1, \mathcal{Q}_2)$ , with the distribution probability  $P_{P_{\text{can}}^{\beta_0}}(\Delta_{\text{exch}}^{\text{excs}, \beta_0} S) = \sum_{\mathcal{Q}_1, \mathcal{Q}_2} \delta(\Delta_{\text{exch}}^{\text{excs}, \beta_0} S - (\beta_1 - \beta_0) \mathcal{Q}_1 - (\beta_2 - \beta_0) \mathcal{Q}_2) P_{P_{\text{can}}^{\beta_0}}(\mathcal{Q}_1, \mathcal{Q}_2; t)$  obeys the symmetry relation at any finite time, which may be referred to as a detailed fluctuation relation,

$$\frac{P_{P_{\text{can}}^{\beta_0}}(\Delta_{\text{exch}}^{\text{excs}, \beta_0} S)}{P_{P_{\text{can}}^{\beta_0}}(-\Delta_{\text{exch}}^{\text{excs}, \beta_0} S)} = e^{-\Delta_{\text{exch}}^{\text{excs}, \beta_0} S}. \quad (4.14)$$

The latter relation itself entails the identity, which may be referred to as an integral fluctuation relation,

$$\langle e^{\Delta_{\text{exch}}^{\text{excs}, \beta_0} S} \rangle_{P_{\text{can}}^{\beta_0}} = 1. \quad (4.15)$$

To our knowledge these two relations have not appeared explicitly in the literature. However they may be seen as another application of the argument first exhibited by Crooks [36] and then Seifert [38, 8] for the entropy production along a stochastic trajectory when the system is in thermal contact with only one heat bath and is driven out of equilibrium by a time-dependent external parameter. (In Crooks' argument the initial configurations for the forward and backward evolutions are distributed with different equilibrium probabilities,  $P_{\text{can}}^{\beta_0}$  and  $P_{\text{can}}^{\beta_f}$ , whereas forward and backward evolutions with the same initial distribution had already been considered in [60, 61]).

## 4.5 Symmetries in protocols starting from a stationary state with a canonical distribution

For some systems, such as the two-spin model studied in paper II, the stationary distribution when the thermostats are at the inverse temperatures  $\beta_1$  and  $\beta_2$  proves to be a canonical distribution at the effective inverse temperature  $\beta_*(\beta_1, \beta_2)$ .

When the system is prepared in a stationary state between two heat baths at the inverse temperatures  $\beta_1^0$  and  $\beta_2^0$  and then put in thermal contact with two thermostats at the inverse temperatures  $\beta_1$  and  $\beta_2$  at time  $t_0 = 0$ , the protocol describes the relaxation from a given stationary state corresponding to  $(\beta_1^0, \beta_2^0)$  to another stationary state corresponding to  $(\beta_1, \beta_2)$ . When the initial stationary state has the canonical distribution at the effective inverse temperature  $\beta_*^0 = \beta_*(\beta_1^0, \beta_2^0)$ , the argument of the previous section can be repeated and the equalities (4.14) and (4.15) still hold with  $\beta_0$  replaced by  $\beta_*^0$  and  $\Delta_{\text{exch}}^{\text{excs}, \beta_0} S$  replaced by

$$\Delta_{\text{exch}}^{\text{excs}, \beta_*^0} S(\mathcal{Q}_1, \mathcal{Q}_2) = (\beta_1 - \beta_*^0) \mathcal{Q}_1 + (\beta_2 - \beta_*^0) \mathcal{Q}_2. \quad (4.16)$$

When the system is already in the stationary state corresponding to the inverse temperatures  $\beta_1$  and  $\beta_2$  at time  $t_0 = 0$ , the equalities (4.14) and (4.15) for  $\Delta_{\text{exch}}^{\text{excs}, \beta_0} S$  still hold with  $\beta_*(\beta_1, \beta_2)$  in place of  $\beta_0$  :

$$\frac{P_{\text{st}} \left( \Delta_{\text{exch}}^{\text{excs}, \beta_*} S \right)}{P_{\text{st}} \left( -\Delta_{\text{exch}}^{\text{excs}, \beta_*} S \right)} = e^{-\Delta_{\text{exch}}^{\text{excs}, \beta_*} S}, \quad (4.17)$$

where the subscript st of the probability is a reminder of the fact that the initial configurations are distributed according to the stationary measure, which is equal to  $P_{\text{can}}^{\beta_*}$  in the present case. Another detailed fluctuation relation involving the forward histories for the original dynamics and the backward histories for the dual reversed dynamics is derived in [62] for the case where the external parameters also vary during the time interval  $]t_0, t]$ ; these considerations are out of the scope of the present paper.

## 5 Long-time symmetries : non-equilibrium steady state with MDB

### 5.1 Fluctuation relations

#### 5.1.1 Fluctuation relation for the cumulated exchange entropy variation

As recalled with some details in section 3, if the Markov matrix is irreducible and the system has a finite number of configurations, then the Perron-Frobenius theorem entails that there exists a single stationary state  $P_{\text{st}}$  and every configuration  $\mathcal{C}$  has a non-zero probability  $P_{\text{st}}(\mathcal{C})$ . Let us call  $P_{\text{st}}^{\min}$  and  $P_{\text{st}}^{\max}$  the minimum and maximum values taken by  $P_{\text{st}}$ . Since  $P_{\text{st}}(\mathcal{Q}_1, \mathcal{Q}_2; t) =$

$\sum_{\mathcal{C}_0, \mathcal{C}_f} P(\mathcal{C}_f | \mathcal{Q}_1, \mathcal{Q}_2; t | \mathcal{C}_0) P_{\text{st}}(\mathcal{C}_0)$ , and  $\frac{P_{\text{st}}^{\min}}{P_{\text{st}}^{\max}} \leq \frac{P_{\text{st}}(\mathcal{C}_f)}{P_{\text{st}}(\mathcal{C}_0)} \leq \frac{P_{\text{st}}^{\max}}{P_{\text{st}}^{\min}}$ , the time-reversal symmetry (4.8) for  $P(\mathcal{C}_f | \mathcal{Q}_1, \mathcal{Q}_2; t | \mathcal{C}_0)$  entails that

$$\frac{P_{\text{st}}^{\min}}{P_{\text{st}}^{\max}} \leq \frac{P_{\text{st}}(\mathcal{Q}_1, \mathcal{Q}_2; t)}{P_{\text{st}}(-\mathcal{Q}_1, -\mathcal{Q}_2; t) e^{-\Delta_{\text{exch}} S(\mathcal{Q}_1, \mathcal{Q}_2)}} \leq \frac{P_{\text{st}}^{\max}}{P_{\text{st}}^{\min}}. \quad (5.1)$$

The distribution probability for the exchange entropy variation  $\Delta_{\text{exch}} S$  can be determined from measurements of heat amounts through the relation

$$P_{\text{st}}(\Delta_{\text{exch}} S; t) = \sum_{\mathcal{Q}_1, \mathcal{Q}_2} \delta(\Delta_{\text{exch}} S - \beta_1 \mathcal{Q}_1 - \beta_2 \mathcal{Q}_2) P_{\text{st}}(\mathcal{Q}_1, \mathcal{Q}_2; t). \quad (5.2)$$

The inequalities (5.1) imply that

$$\frac{P_{\text{st}}^{\min}}{P_{\text{st}}^{\max}} \leq \frac{P_{\text{st}}(\Delta_{\text{exch}} S; t)}{P_{\text{st}}(-\Delta_{\text{exch}} S; t) e^{-\Delta_{\text{exch}} S}} \leq \frac{P_{\text{st}}^{\max}}{P_{\text{st}}^{\min}}. \quad (5.3)$$

In the long-time limit the system reaches its non-equilibrium stationary state exponentially fast, so the existence and value of a large deviation function for the cumulated current  $\mathcal{J} \equiv \Delta_{\text{exch}} S/t$  are not expected to depend on the initial distribution. Appendix E contains several definitions of large deviation functions. Subsection E.3 contains a proof that, whatever the definition, a relation like (5.3) implies that, if  $\Delta_{\text{exch}} S$  has a large deviation function  $f_{\Delta_{\text{exch}} S}(\mathcal{J})$  – under  $P_{\text{st}}$  but then also under any initial probability distribution – then

$$f_{\Delta_{\text{exch}} S}(\mathcal{J}) - f_{\Delta_{\text{exch}} S}(-\mathcal{J}) = -\mathcal{J}. \quad (5.4)$$

Our derivation of this fluctuation relation is close to the argument given in Ref.[41] for the large deviation of the cumulated heat current (see (5.9)). It relies on the MDB obeyed by the transition rates. Since the opposite of the exchange entropy variation is the specific form which the Lebowitz-Spohn action functional [1] takes in the presence of MDB, the fluctuation relation (5.4) is in fact a special case of the fluctuation relation satisfied by the action functional for a system with a finite number of configurations under the assumption (1.2) of reversibility for configuration jumps and the assumption (1.1) that the Markov matrix is irreducible, without the extra assumption of MDB (1.3).

### 5.1.2 Constraints from the bound upon the system energy

In a system with a finite number of configurations,  $\mathcal{Q}_1 + \mathcal{Q}_2 = \mathcal{E}(\mathcal{C}_f) - \mathcal{E}(\mathcal{C}_0)$  is bounded and this entails several properties upon the large deviation functions of the cumulated currents  $\mathcal{J}_1 = \mathcal{Q}_1/t$  and  $\mathcal{J}_2 = \mathcal{Q}_2/t$ .

When  $\mathcal{Q}_1 + \mathcal{Q}_2$  is bounded, the consequences for the cumulated heat currents are conveniently investigated if one considers the couple of variables  $(\mathcal{Q}_1^d, \mathcal{Q}_2)$  where  $\mathcal{Q}_1^d = -\mathcal{Q}_1$  is the heat amount dissipated towards the thermal bath 1. The fact that the difference  $\mathcal{Q}_1^d - \mathcal{Q}_2$  is bounded means that there exists some (time-independent) constant  $M > 0$  such that

$$|\mathcal{Q}_1^d - \mathcal{Q}_2| < M. \quad (5.5)$$

The first straightforward consequence upon the cumulated heat currents  $\mathcal{J}_1^d = \mathcal{Q}_1^d/t$  and  $\mathcal{J}_2 = \mathcal{Q}_2/t$  is that  $\lim_{t \rightarrow +\infty} \langle \mathcal{J}_1^d \rangle = \lim_{t \rightarrow +\infty} \langle \mathcal{J}_2 \rangle \equiv J$ , namely

$$\lim_{t \rightarrow +\infty} \frac{-\langle \mathcal{Q}_1 \rangle}{t} = \lim_{t \rightarrow +\infty} \frac{\langle \mathcal{Q}_2 \rangle}{t} = J. \quad (5.6)$$

The second consequence of the fact that  $\mathcal{Q}_1^d - \mathcal{Q}_2$  is sub-extensive is that, according to (E.30),  $\mathcal{Q}_1^d$  and  $\mathcal{Q}_2$  have the same large deviation function,  $f_{\mathcal{Q}_1^d}(\mathcal{J}) = f_{\mathcal{Q}_2}(\mathcal{J})$ , namely

$$f_{\mathcal{Q}_1}(\mathcal{J}) = f_{\mathcal{Q}_2}(-\mathcal{J}). \quad (5.7)$$

Similarly the difference between  $\Delta_{\text{exch}}S$  and  $-(\beta_1 - \beta_2)Q_2$ , which is equal to  $\beta_1[\mathcal{E}(\mathcal{C}_f) - \mathcal{E}(\mathcal{C}_f)]$ , is bounded. A fortiori  $\Delta_{\text{exch}}S + (\beta_1 - \beta_2)Q_2$  is sub-extensive, so that according to (E.30)  $f_{\Delta_{\text{exch}}S}(\mathcal{J}) = f_{-(\beta_1 - \beta_2)Q_2}(\mathcal{J})$ , namely

$$f_{\Delta_{\text{exch}}S}(\mathcal{J}) = f_{Q_2}\left(-\frac{\mathcal{J}}{\beta_1 - \beta_2}\right). \quad (5.8)$$

As a consequence, the fact that the exchange entropy variation  $\Delta_{\text{exch}}S$  obeys the fluctuation relation (5.4) is equivalent to the fact that the cumulated heat current received from the heat bath 2 obeys the fluctuation relation

$$f_{Q_2}(\mathcal{J}) - f_{Q_2}(-\mathcal{J}) = (\beta_1 - \beta_2)\mathcal{J}. \quad (5.9)$$

The latter equation is the long-time limit of the relation first exhibited by Jarzynski and Wojcik [42] for the thermal contact between two Hamiltonian systems initially prepared at different inverse temperatures  $\beta_1$  and  $\beta_2$  and set in contact at time  $t_0 = 0$ , with the implicit assumption that the energy variation of the system which ensures the thermal contact is negligible with respect to the heat quantity which goes through it. The fluctuation relation (5.9) is also derived in the framework of the master equation approach in Refs.[40, 41].

We notice that, since the function  $\Delta_{\text{exch}}^{\text{excs}, \beta_0}S(Q_1, Q_2)$  defined in (4.12) is equal to  $\Delta_{\text{exch}}S$  plus a term  $-\beta_0(Q_1 + Q_2)$  which is bounded by virtue of energy conservation, the large deviation functions for  $\Delta_{\text{exch}}^{\text{excs}, \beta_0}S$  and  $\Delta_{\text{exch}}S$  coincide (see (E.30)). As a consequence, (5.4) entails that  $\Delta_{\text{exch}}^{\text{excs}, \beta_0}(Q_1, Q_2)$  obeys the fluctuation relation  $f_{\Delta_{\text{exch}}^{\text{excs}, \beta_0}}(\mathcal{J}) - f_{\Delta_{\text{exch}}^{\text{excs}, \beta_0}}(-\mathcal{J}) = -\mathcal{J}$ .

The relation (5.9) can be written in a more generic form by the following argument. The fact that the difference between  $\Delta_{\text{exch}}S$  and  $-(\beta_1 - \beta_2)Q_2$  is bounded yields the following relation between the infinite-time mean values of the corresponding cumulated currents,

$$\lim_{t \rightarrow +\infty} \frac{\langle \Delta_{\text{exch}}S(t) \rangle}{t} = -(\beta_1 - \beta_2) \lim_{t \rightarrow +\infty} \frac{\langle Q_2(t) \rangle}{t}. \quad (5.10)$$

Since the infinite-time limit of the mean value of a cumulated current  $\mathcal{J}_t$  measured during the interval  $[0, t]$  is equal to the mean value of the corresponding instantaneous current in the stationary state, namely

$$\lim_{t \rightarrow \infty} \langle \mathcal{J}_t \rangle = \langle j \rangle_{\text{st}} \equiv J, \quad (5.11)$$

we get  $\lim_{t \rightarrow +\infty} \langle \Delta_{\text{exch}}S \rangle / t = \langle j_{\delta_{\text{exch}}S} \rangle_{\text{st}} = d_{\text{exch}}S/dt|_{\text{st}}$  as well as  $\lim_{t \rightarrow +\infty} \langle Q_2 \rangle / t = \langle j_2 \rangle_{\text{st}}$  (with the instantaneous current definitions (3.18) and (3.17) respectively). With these identifications the comparison of (5.10) with the property (1.4) (which exhibits the analogy with the thermodynamics of irreversible processes) shows that  $\beta_1 - \beta_2$  in (5.10) is to be interpreted as the thermodynamic force  $\mathcal{F}$ . Therefore the fact that the fluctuation relation (5.9) arises from the boundedness of the difference between  $\Delta_{\text{exch}}S$  and  $-(\beta_1 - \beta_2)Q_2$ , as the relation between the mean values (5.10) does, implies that the fluctuation relation (5.9) for  $f_{Q_2}$  is a special case of the more generic fluctuation relation

$$f(\mathcal{J}; \mathcal{F}) - f(-\mathcal{J}; \mathcal{F}) = \mathcal{F}\mathcal{J}, \quad (5.12)$$

where  $\mathcal{F}$  is the thermodynamic force which appears in the stationary exchange entropy flow  $d_{\text{exch}}S/dt|_{\text{st}} = -\mathcal{F}J$ .

### 5.1.3 The Gärtner-Ellis theorem and some of its consequences

As exemplified in the previous subsection, large deviation functions may depend on auxiliary parameters. For certain questions, they are simply spectators. This is the case in the following discussion, so we do not mention possible auxiliary parameters explicitly. However, we nevertheless write all derivatives as partial derivatives.

In this short subsection, we introduce an important tool to study the existence and properties of a large deviation function  $f(\mathcal{J})$ : the infinite-time limit of the generating function for the

cumulants of  $X_t = t\mathcal{J}_t$  per unit time, namely  $\alpha(\lambda) \equiv \lim_{t \rightarrow +\infty} (1/t) \ln \langle e^{\lambda X_t} \rangle$ . This function, if it exists, is automatically convex.

According to a simplified version of the Gärtner-Ellis theorem (see e.g. the review for physicists [63] or the mathematical point of view [64]), if  $\alpha(\lambda)$  exists and is differentiable for all  $\lambda$  in  $\mathbb{R}$ , then the large deviation function  $f$  of the current  $\mathcal{J}$  exists and it can be calculated as the Legendre-Fenchel transform of  $\alpha(\lambda)$ , namely, with the signs chosen in the definitions used in the present paper,

$$f(\mathcal{J}) = \min_{\lambda \in \mathbb{R}} \{ \alpha(\lambda) - \lambda \mathcal{J} \}. \quad (5.13)$$

If  $\alpha(\lambda)$  is stricly convex and continuously differentiable, then for each  $\mathcal{J}$  the minimum is achieved for a single value of  $\lambda$ , which is the unique solution  $\lambda_c$  of  $\frac{\partial \alpha}{\partial \lambda} = \mathcal{J}$  i.e. the Legendre-Fenchel transform reduces to the usual Legendre transform, and the duality relation,  $\lambda_c(\mathcal{J}) = -\frac{\partial f}{\partial \mathcal{J}}$  holds, i.e.

$$\left. \frac{\partial \alpha}{\partial \lambda} \right|_{\lambda = -\frac{\partial f}{\partial \mathcal{J}}} = \mathcal{J}. \quad (5.14)$$

If moreover  $\alpha(\lambda)$  is differentiable twice, taking the derivative of this relation with respect to  $\mathcal{J}$  one obtains

$$\left. \frac{\partial^2 f}{(\partial \mathcal{J})^2} \frac{\partial^2 \alpha}{(\partial \lambda)^2} \right|_{\lambda = \lambda_c(\mathcal{J})} = -1. \quad (5.15)$$

From the fundamental properties of any large deviation function,  $f(\mathcal{J})$  is maximum at  $\mathcal{J} = J$  defined in (5.11), so by construction  $\lambda_c(J) = 0$ . But derivatives of  $\alpha(\lambda)$  at  $\lambda = 0$  are related to cumulants of  $X_t$  at large  $t$ . For instance

$$\left. \frac{\partial^2 \alpha}{(\partial \lambda)^2} \right|_{\lambda=0} = \lim_{t \rightarrow +\infty} \frac{\langle X_t^2 \rangle_{st} - \langle X_t \rangle_{st}^2}{t}, \quad (5.16)$$

and one gets:

$$\left. \frac{\partial^2 f}{(\partial \mathcal{J})^2} \right|_{\mathcal{J}=J} = - \left[ \lim_{t \rightarrow +\infty} \frac{\langle X_t^2 \rangle_{st} - \langle X_t \rangle_{st}^2}{t} \right]^{-1}. \quad (5.17)$$

We now apply these considerations to linear response.

#### 5.1.4 Linear response far from equilibrium

In the case at hand, the large deviation function  $f$  depends on other parameters. The one relevant for the discussion of linear response is the thermodynamic force  $\mathcal{F}$ , and we write  $f(\mathcal{J}; \mathcal{F})$ . In general  $f$  depends on other variables that we do not mention explicitly, and which are supposed to be kept constant whenever a partial derivative is taken in the sequel.

We start from the property recalled in the previous subsection:

$$\left. \frac{\partial f(\mathcal{J}; \mathcal{F})}{\partial \mathcal{J}} \right|_{\mathcal{J}=J} = 0. \quad (5.18)$$

Taking the derivative with respect to  $\mathcal{F}$  leads to

$$\left. \frac{\partial J}{\partial \mathcal{F}} \frac{\partial^2 f(\mathcal{J}; \mathcal{F})}{(\partial \mathcal{J})^2} \right|_{\mathcal{J}=J} = - \left. \frac{\partial^2 f(\mathcal{J}; \mathcal{F})}{\partial \mathcal{F} \partial \mathcal{J}} \right|_{\mathcal{J}=J}. \quad (5.19)$$

According to (5.17),  $\left. \frac{\partial^2 f(\mathcal{J}; \mathcal{F})}{(\partial \mathcal{J})^2} \right|_{\mathcal{J}=J}$  can be expressed in terms of the second cumulant of  $X_t$  in the infinite-time limit, and we can rewrite the equality (5.19) as:

$$\frac{\partial J}{\partial \mathcal{F}} = \left. \frac{\partial^2 f(\mathcal{J}; \mathcal{F})}{\partial \mathcal{F} \partial \mathcal{J}} \right|_{\mathcal{J}=J} \times \lim_{t \rightarrow +\infty} \frac{\langle X_t^2 \rangle_{st} - \langle X_t \rangle_{st}^2}{t}. \quad (5.20)$$

This is the generic expression of the linear response in the non-equilibrium state far from equilibrium. In the case of thermal contact  $X_t$  is the cumulated heat  $Q_2(t)$  received from the heat bath 2.



### 5.1.5 Green-Kubo relation as a consequence of the fluctuation relation for the large deviation function

When the large deviation function obeys the fluctuation relation (5.12) arising from the MDB, successive partial derivatives of the fluctuation relation entail that

$$\frac{\partial^2 f(\mathcal{J}; \mathcal{F})}{\partial \mathcal{F} \partial \mathcal{J}} + \frac{\partial^2 f(\mathcal{J}'; \mathcal{F})}{\partial \mathcal{F} \partial \mathcal{J}'} \Big|_{\mathcal{J}' = -\mathcal{J}} = 1. \quad (5.21)$$

Then the relation (5.21) for  $\mathcal{J} = 0$  yields  $\frac{\partial^2 f(\mathcal{J}; \mathcal{F})}{\partial \mathcal{F} \partial \mathcal{J}} \Big|_{\mathcal{J}=0} = \frac{1}{2}$  and, since the stationary state with  $\mathcal{F} = 0$  is in fact the equilibrium state,

$$\frac{\partial J}{\partial \mathcal{F}} \Big|_{\mathcal{F}=0} = \frac{1}{2} \times \lim_{t \rightarrow +\infty} \frac{\langle X_t^2 \rangle_{\text{eq}} - \langle X_t \rangle_{\text{eq}}^2}{t}. \quad (5.22)$$

In the following the latter fluctuation-dissipation relation will be referred to as the Green-Kubo relation. The precise terminology has been pointed out in Ref.[51]. When the equilibrium fluctuations are written in terms of a second-order cumulant, namely the mean value of the squared Helfand moment  $X_t - \langle X_t \rangle$ , as it is the case for Einstein relation, the fluctuation-dissipation relations are called Einstein-Helfand formulae [13, 65]. When the equilibrium fluctuations are written in terms of the time-correlation function of the instantaneous current, they are known as the Green-Kubo relations [66, 67, 68, 69] or the Yamamoto-Zwanzig formulae in the context of chemical relations [70, 71].

The Green-Kubo relation (5.22) can be rephrased in terms of the Onsager coefficient  $L$  defined in the framework of thermodynamics of irreversible processes near equilibrium as

$$L \equiv \frac{\partial J}{\partial \mathcal{F}} \Big|_{\mathcal{F}=0} = \lim_{\mathcal{F} \rightarrow 0} \frac{J}{\mathcal{F}}, \quad (5.23)$$

where the last equality is valid when there is only one nonzero mean current in the stationary state, in which case this single current  $J$  vanishes as  $\mathcal{F}$  goes to zero.

In the case of thermal contact, the limit  $\mathcal{F} \rightarrow 0$  in (5.23) can be stated more precisely according to the following argument. If  $\beta_1 = \beta_2$  the stationary state is the equilibrium state and  $J(\beta_1, \beta_2)$  becomes  $J(\beta, \beta) = \langle j_2 \rangle_{\text{eq}} = 0$  (see the comment after (3.17)). As a consequence  $J(\beta + d\beta, \beta + d\beta) - J(\beta, \beta) = 0$  for any infinitesimal  $d\beta$ , namely

$$\frac{\partial J}{\partial \beta_1} \Big|_{\beta_1=\beta_2=\beta} = - \frac{\partial J}{\partial \beta_2} \Big|_{\beta_1=\beta_2=\beta}, \quad (5.24)$$

so that

$$J(\beta_1, \beta_2) \underset{\beta_1, \beta_2 \rightarrow \beta}{\sim} (\beta_1 - \beta_2) \frac{\partial J}{\partial \beta_1} \Big|_{\beta_1=\beta_2=\beta} \quad (5.25)$$

independently of the way  $\beta_1$  and  $\beta_2$  go to  $\beta$ . Subsequently the linear response coefficient  $L$  defined in (5.23) also reads

$$L = \frac{\partial J}{\partial \beta_1} \Big|_{\beta_1=\beta_2=\beta}, \quad (5.26)$$

while the Green-Kubo relation (5.22) can also be stated as

$$\lim_{\beta_1, \beta_2 \rightarrow \beta} \frac{J(\beta_1, \beta_2)}{\beta_1 - \beta_2} = \frac{1}{2} \lim_{t \rightarrow +\infty} \frac{\langle Q_2^2(t) \rangle_{\text{eq}}^\beta - (\langle Q_2(t) \rangle_{\text{eq}}^\beta)^2}{t}. \quad (5.27)$$

where  $J(\beta_1, \beta_2) = \langle j_2 \rangle_{\text{st}}^{(\beta_1, \beta_2)}$  and the limit is independent of the way  $\beta_1 - \beta_2$  vanishes. Eventually in the case of thermal contact the Green-Kubo relation relates the proportionality coefficient between the stationary heat current and the difference between the bath inverse temperatures when the system is weakly out of equilibrium to the fluctuations of the heat amount received from one thermal bath in the equilibrium situation where both thermostats are at the same temperature.

## 5.2 Symmetry of the generating function for cumulants per unit time under MDB

Contrarily to the case of the Green-Kubo relation (5.22) between the mean current and the equilibrium infinite-time second cumulant per unit time, relations between higher cumulants per unit time cannot be readily obtained from the fluctuation relation (5.12). In order to obtain such relations one has to resort to the cumulant generating function.

### 5.2.1 Infinite-time cumulants per unit time for $\mathcal{Q}_1$ and $\mathcal{Q}_2$

The  $k$ th cumulant  $\kappa_{\mathcal{Q}_a}^{[p]}$  for the heat amount  $\mathcal{Q}_a$  received from the bath  $a$  or the joint cumulant  $\kappa_{\mathcal{Q}_a, \mathcal{Q}_b}^{[p, q]}$  for the heat amounts  $\mathcal{Q}_a$  and  $\mathcal{Q}_b$  can be computed from the characteristic function

$$\langle e^{\lambda_1 \mathcal{Q}_1(t) + \lambda_2 \mathcal{Q}_2(t)} \rangle = \sum_{\mathcal{Q}_1} \sum_{\mathcal{Q}_2} e^{\lambda_1 \mathcal{Q}_1 + \lambda_2 \mathcal{Q}_2} \sum_{\mathcal{C}_f} \sum_{\mathcal{C}_0} P(\mathcal{C}_f | \mathcal{Q}_1, \mathcal{Q}_2; t | \mathcal{C}_0) P(\mathcal{C}_0, t_0 = 0) \quad (5.28)$$

through the following derivatives

$$\kappa_{\mathcal{Q}_a}^{[p]}(t) = \left. \frac{\partial^p \ln \langle e^{\lambda_1 \mathcal{Q}_1(t) + \lambda_2 \mathcal{Q}_2(t)} \rangle}{\partial \lambda_a^p} \right|_{\lambda_1 = \lambda_2 = 0} \quad \text{for } a = \{1, 2\} \quad (5.29)$$

$$\kappa_{\mathcal{Q}_1, \mathcal{Q}_2}^{[p, q]}(t) = \left. \frac{\partial^{p+q} \ln \langle e^{\lambda_1 \mathcal{Q}_1(t) + \lambda_2 \mathcal{Q}_2(t)} \rangle}{\partial \lambda_1^p \partial \lambda_2^q} \right|_{\lambda_1 = \lambda_2 = 0}. \quad (5.30)$$

For a Markov process, the leading long-time behaviors of these cumulants are proportional to the time  $t$  elapsed from the beginning of the measurements. The asymptotic behavior of the cumulants per unit time are given by the derivatives of

$$\alpha_{1,2}(\lambda_1, \lambda_2) \equiv \lim_{t \rightarrow +\infty} \frac{1}{t} \ln \langle e^{\lambda_1 \mathcal{Q}_1(t) + \lambda_2 \mathcal{Q}_2(t)} \rangle \quad (5.31)$$

with respect to  $\lambda_1$  and  $\lambda_2$  at  $\lambda_1 = \lambda_2 = 0$ . In other words,

$$\lim_{t \rightarrow +\infty} \frac{\kappa_{\mathcal{Q}_1, \mathcal{Q}_2}^{[p, q]}(t)}{t} = \left. \frac{\partial^{p+q} \alpha_{1,2}}{\partial \lambda_1^p \partial \lambda_2^q} \right|_{\lambda_1 = \lambda_2 = 0}. \quad (5.32)$$

Similarly

$$\lim_{t \rightarrow +\infty} \frac{\kappa_{\mathcal{Q}_a}^{[p]}(t)}{t} = \left. \frac{\partial^p \alpha_a}{\partial \lambda_a^p} \right|_{\lambda_a = 0} \quad \text{for } a = \{1, 2\} \quad (5.33)$$

where

$$\alpha_a(\lambda) \equiv \lim_{t \rightarrow +\infty} \frac{1}{t} \ln \langle e^{\lambda \mathcal{Q}_a(t)} \rangle. \quad (5.34)$$

The generating function for infinite-time cumulants per unit time,  $\alpha_a(\lambda)$ , may also be referred to as the scaled cumulant generating function.

We notice that, since for any cumulant  $\lim_{t \rightarrow \infty} \kappa_{\mathcal{Q}_2}^{[p]}(t)/t$  is finite, under some technical conditions, the probability distribution of the variable  $[\mathcal{Q}_2(t) - \langle \mathcal{Q}_2(t) \rangle] / \sqrt{t}$  becomes Gaussian in the long-time limit. Indeed the logarithm of the characteristic function for the variable  $Y_2(t) = [\mathcal{Q}_2(t) - \langle \mathcal{Q}_2(t) \rangle] / \sqrt{t}$  reads  $\ln \langle e^{\lambda Y_2(t)} \rangle = \sum_{p=2}^{+\infty} (1/p!) (\lambda/\sqrt{t})^p \kappa_{\mathcal{Q}_2}^{[p]}(t)$ . If the sum and the  $t \rightarrow +\infty$  limit can be interchanged,  $\ln \langle e^{\lambda Y_2(t)} \rangle$  becomes proportional to  $\lambda^2$  in the limit where  $t$  goes to infinity : only the second cumulant of  $Y_2(t)$  survives in the long-time limit.

### 5.2.2 Constraints from the bound upon the system energy

In the case of a system with a finite number of configurations,  $\mathcal{Q}_1 + \mathcal{Q}_2 = \mathcal{E}(\mathcal{C}_f) - \mathcal{E}(\mathcal{C}_0)$  is restricted to some interval  $[-|\Delta\mathcal{E}|_{\max}, +|\Delta\mathcal{E}|_{\max}]$ , and the definition (5.28) entails the inequalities

$$e^{-\lambda_1 |\Delta\mathcal{E}|_{\max}} \leq \frac{\langle e^{\lambda_1 \mathcal{Q}_1(t) + \lambda_2 \mathcal{Q}_2(t)} \rangle}{\langle e^{(\lambda_2 - \lambda_1) \mathcal{Q}_2(t)} \rangle} \leq e^{\lambda_1 |\Delta\mathcal{E}|_{\max}}. \quad (5.35)$$

As a consequence,

$$\alpha_{1,2}(\lambda_1, \lambda_2) = \alpha_a(\lambda_a - \lambda_b) \quad \text{with } \{a, b\} = \{1, 2\}. \quad (5.36)$$

The relation (5.36) can be rewritten as

$$\alpha_{1,2}(\lambda_1, \lambda_2) = \alpha_2(\lambda_2 - \lambda_1) \quad \text{and} \quad \alpha_1(\lambda) = \alpha_2(-\lambda). \quad (5.37)$$

The specific dependence of  $\alpha_{12}(\lambda_1, \lambda_2)$  upon  $\lambda_2 - \lambda_1$  together with the generic formulæ (5.32)–(5.33) imply the following relations between the infinite-time cumulants per unit time,

$$\lim_{t \rightarrow \infty} \frac{\kappa_{\mathcal{Q}_1}^{[p]}}{t} = (-1)^p \lim_{t \rightarrow \infty} \frac{\kappa_{\mathcal{Q}_2}^{[p]}}{t} \quad (5.38)$$

and

$$\lim_{t \rightarrow \infty} \frac{\kappa_{\mathcal{Q}_1, \mathcal{Q}_2}^{[p, q]}}{t} = (-1)^p \lim_{t \rightarrow \infty} \frac{\kappa_{\mathcal{Q}_2}^{[p+q]}}{t}. \quad (5.39)$$

### 5.2.3 MDB and symmetry of the generating function for infinite-time cumulants per unit time

The modified detailed balance entails the time-reversal symmetry (4.8) for  $P(\mathcal{C}_f | \mathcal{Q}_1, \mathcal{Q}_2; t | \mathcal{C}_0)$  at finite time. Henceforth, according to its definition (5.28), the characteristic function can be rewritten as

$$\langle e^{\lambda_1 \mathcal{Q}_1(t) + \lambda_2 \mathcal{Q}_2(t)} \rangle = \sum_{\mathcal{Q}_1} \sum_{\mathcal{Q}_2} e^{(\beta_1 - \lambda_1) \mathcal{Q}_1 + (\beta_2 - \lambda_2) \mathcal{Q}_2} \sum_{\mathcal{C}_f} \sum_{\mathcal{C}_0} P(\mathcal{C}_0 | \mathcal{Q}_1, \mathcal{Q}_2; t | \mathcal{C}_f) P(\mathcal{C}_0, t_0 = 0). \quad (5.40)$$

When the Markov matrix is irreducible and the system has a finite number of configurations, the Perron-Frobenius theorem entails that there exists a single stationary state  $P_{\text{st}}$ , and every configuration  $\mathcal{C}$  has a non-zero probability  $P_{\text{st}}(\mathcal{C})$ . Henceforth, in the case where the initial distribution is the stationary one, the relation (5.40) leads to the inequalities

$$\frac{P_{\text{st}}^{\min}}{P_{\text{st}}^{\max}} \leq \frac{\langle e^{(\beta_1 - \lambda_1) \mathcal{Q}_1 + (\beta_2 - \lambda_2) \mathcal{Q}_2} \rangle_{\text{st}}}{\langle e^{\lambda_1 \mathcal{Q}_1 + \lambda_2 \mathcal{Q}_2} \rangle_{\text{st}}} \leq \frac{P_{\text{st}}^{\max}}{P_{\text{st}}^{\min}}. \quad (5.41)$$

(We recall that  $\langle \dots \rangle_{\text{st}}$  denotes an average when the initial configurations are distributed according to the stationary measure  $P_{\text{st}}(\mathcal{C})$ , the maximum and minimum values of which are  $P_{\text{st}}^{\max}$  and  $P_{\text{st}}^{\min}$  respectively (see (5.1)). These inequalities entail that the generating function for the infinite-time limits of the joint cumulants per unit time defined in (5.31) obeys the symmetry

$$\alpha_{12}(\lambda_1, \lambda_2) = \alpha_{12}(\beta_1 - \lambda_1, \beta_2 - \lambda_2). \quad (5.42)$$

We notice that the symmetry (5.42) can also be derived by considering the evolution of the Laplace transform of  $P(\mathcal{Q}_1, \mathcal{Q}_2; t)$ . With the notations of paper II [45],  $\langle e^{\lambda_1 \mathcal{Q}_1(t) + \lambda_2 \mathcal{Q}_2(t)} \rangle_{\text{st}} = \sum_{\mathcal{C}_f} \sum_{\mathcal{C}_0} (\mathcal{C}_f | \hat{\mathbb{U}}(e^{\lambda_1}, e^{\lambda_2}; t) | \mathcal{C}_0) P_{\text{st}}(\mathcal{C}_0)$  and  $\alpha_{12}(\lambda_1, \lambda_2)$  coincides with the largest eigenvalue of the operator  $\hat{\mathbb{A}}(\lambda_1, \lambda_2)$  which rules the evolution of  $\hat{\mathbb{U}}(e^{\lambda_1}, e^{\lambda_2}; t)$  according to  $d\hat{\mathbb{U}}/dt = \hat{\mathbb{A}}\hat{\mathbb{U}}$ . The MDB implies that the operator  $\hat{\mathbb{A}}(\lambda_1, \lambda_2)$  obeys the symmetry  $\hat{\mathbb{A}}(\lambda_1, \lambda_2) = \tilde{\mathbb{A}}^T(\beta_1 - \lambda_1, \beta_2 - \lambda_2)$ , where  $\tilde{\mathbb{A}}^T$  denotes the transposed matrix of  $\hat{\mathbb{A}}$  in the configuration basis. Then, by using the Perron-Frobenius theorem, one can prove that the largest eigenvalue of  $\hat{\mathbb{A}}(\lambda_1, \lambda_2)$  satisfies the symmetry

relation (5.42) (see the argument in [1] where an analogous symmetry is exhibited and then used for the derivation of the long-time fluctuation relation obeyed by the action functional defined in the comment after (4.5)-(4.6)).

Moreover, since the system has a finite number of configurations,  $\alpha_{12}(\lambda_1, \lambda_2) = \alpha_a(\lambda_a - \lambda_b)$  by virtue of (5.36), and the symmetry property (5.42) of  $\alpha_{12}(\lambda_1, \lambda_2)$  becomes

$$\alpha_a(\lambda) = \alpha_a(\beta_a - \beta_b - \lambda) \quad \text{for } \{a, b\} = \{1, 2\}. \quad (5.43)$$

We can apply the Gärtner-Ellis theorem (see subsubsection 5.1.3 and references there): if  $\alpha_2(\lambda)$  exists and is differentiable for all  $\lambda$  in  $\mathbb{R}$ , then the large deviation function of the current  $\mathcal{J} = \mathcal{Q}_2/t$  exists and it can be calculated as the Legendre-Fenchel transform of  $\alpha_2(\lambda)$ , namely, with the signs chosen in the definitions used in the present paper,

$$f_{\mathcal{Q}_2}(\mathcal{J}) = \min_{\lambda \in \mathbb{R}} \{\alpha_2(\lambda) - \lambda \mathcal{J}\}. \quad (5.44)$$

Then, since  $\alpha_2(\lambda)$  obeys the symmetry (5.43),  $f_{\mathcal{Q}_2}(\mathcal{J})$  obeys the fluctuation relation (5.9). We notice that we can similarly retrieve the fluctuation relation obeyed by  $f_{\Delta_{\text{exch}}S}$ . Indeed,  $\alpha_{\Delta_{\text{exch}}S}(\lambda) \equiv \lim_{t \rightarrow +\infty} (1/t) \ln \langle e^{\lambda \Delta_{\text{exch}}S(t)} \rangle$  is equal to  $\alpha_{12}(\lambda\beta_1, \lambda\beta_2)$  with  $\alpha_{12}(\lambda_1, \lambda_2)$  defined in (5.31), and, according to (5.36),  $\alpha_{\Delta_{\text{exch}}S}(\lambda) = \alpha_a(\lambda(\beta_a - \beta_b))$ . Therefore the symmetry (5.43) implies that

$$\alpha_{\Delta_{\text{exch}}S}(1 - \lambda) = \alpha_{\Delta_{\text{exch}}S}(\lambda) \quad (5.45)$$

Under the above assumptions, the large deviation function of the exchange entropy cumulated current can be computed as the Legendre-Fenchel transform of  $\alpha_{\Delta_{\text{exch}}S}$ , namely  $f_{\Delta_{\text{exch}}S}(\mathcal{J}) = \min_{\lambda \in \mathbb{R}} \{\alpha_{\Delta_{\text{exch}}S}(\lambda) - \lambda \mathcal{J}\}$ , the latter symmetry allows to retrieve the fluctuation relation (5.4).

Another consequence of the symmetry (5.43) is that the Green-Kubo relation near equilibrium, which can be derived from the fluctuation relation (5.9) for  $f_{\mathcal{Q}_2}(\mathcal{J})$  as shown in section 5.1.5, can equivalently be directly derived from the symmetry (5.43) of its Legendre-Fenchel transform  $\alpha_2(\lambda)$ . This can be checked as follows. Relation (5.43) can be rewritten for  $a = 2$  as

$$\alpha_2(\lambda; \beta_1, \beta_2) = \alpha_2(-\mathcal{F} - \lambda; \beta_1, \beta_2) \quad (5.46)$$

where  $\mathcal{F} = \beta_1 - \beta_2$  is the thermodynamic force. The Green-Kubo relation involves the variation of the mean current  $J(\beta_1, \beta_2) \equiv \frac{\partial \alpha_2}{\partial \lambda} \Big|_{(\lambda=0; \beta_1, \beta_2)}$  with respect to  $\mathcal{F}$ , so we need to compute mixed derivatives of  $\alpha_2$  with respect to  $\lambda$  and  $\mathcal{F}$ . Notice that in the identity (5.46), the second and the third arguments are untouched. So if we make any invertible ( $\lambda$ -independent) change of variables  $(\beta_1, \beta_2) \leftrightarrow (\mathcal{F}, \rho)$  and set  $\alpha_2(\lambda; \beta_1, \beta_2) \equiv \alpha(\lambda; \mathcal{F}, \rho)$ , we have the following identity

$$\alpha(\lambda; \mathcal{F}, \rho) = \alpha(-\mathcal{F} - \lambda; \mathcal{F}, \rho) \quad (5.47)$$

where the third variable  $\rho$  is purely a spectator. Keeping it fixed, we get from the symmetry (5.47):

$$\frac{\partial^2 \alpha}{\partial \mathcal{F} \partial \lambda} \Big|_{(\lambda; \mathcal{F})} = - \frac{\partial^2 \alpha}{\partial \mathcal{F} \partial \lambda} \Big|_{(-\mathcal{F} - \lambda; \mathcal{F})} + \frac{\partial^2 \alpha}{\partial \lambda^2} \Big|_{(-\mathcal{F} - \lambda; \mathcal{F})}. \quad (5.48)$$

Taking this relation at  $\lambda = \mathcal{F} = 0$ , we obtain

$$2 \frac{\partial^2 \alpha}{\partial \mathcal{F} \partial \lambda} \Big|_{(\lambda=0; \mathcal{F}=0)} = \frac{\partial^2 \alpha}{\partial \lambda^2} \Big|_{(\lambda=0; \mathcal{F}=0)}. \quad (5.49)$$

From now on, the discussion parallels the one in section 5.1.5. We repeat the argument in a slightly different form to stress again the slight subtlety involved in variations with respect to  $\mathcal{F}$  in the context of thermal contact. The condition  $\mathcal{F} = 0$  means  $\beta_1 = \beta_2 \equiv \beta$ . So the right-hand side of (5.49) is simply  $\frac{\partial^2 \alpha}{\partial \lambda^2} \Big|_{(\lambda=0; \mathcal{F}=0)} = \frac{\partial^2 \alpha_2}{\partial \lambda^2} \Big|_{(\lambda=0; \beta_1=\beta_2=\beta)}$ . The left-hand side of (5.49) can be

dealt with as follows. The change of variables  $(\beta_1, \beta_2) \leftrightarrow (\mathcal{F}, \rho)$  is  $\lambda$ -independent, so we can set  $\lambda = 0$  once the derivative with respect to  $\lambda$  is taken

$$\left. \frac{\partial \alpha}{\partial \lambda} \right|_{(\lambda=0; \mathcal{F}, \rho)} = \left. \frac{\partial \alpha_2}{\partial \lambda} \right|_{(\lambda=0; \beta_1, \beta_2)} = J(\beta_1, \beta_2). \quad (5.50)$$

Of course, taking the partial derivative of  $J(\beta_1, \beta_2)$  with respect to  $\mathcal{F}$  is ambiguous, as it depends on the choice of the variable  $\rho$ . However,  $\left. \frac{\partial J}{\partial \mathcal{F}} \right|_{\mathcal{F}=0}$  has an intrinsic meaning, independent of the choice of the variable  $\rho$ . Indeed, by construction  $J(\beta, \beta) = 0$  (equilibrium) so  $\left. \frac{\partial J}{\partial \beta_1} \right|_{\beta_1=\beta_2=\beta} + \left. \frac{\partial J}{\partial \beta_2} \right|_{\beta_1=\beta_2=\beta} = 0$ . Moreover, from  $\mathcal{F} = \beta_1 - \beta_2$  we get  $1 = \frac{\partial \beta_1}{\partial \mathcal{F}} - \frac{\partial \beta_2}{\partial \mathcal{F}}$ . Now  $\frac{\partial J}{\partial \mathcal{F}} = \frac{\partial J}{\partial \beta_1} \frac{\partial \beta_1}{\partial \mathcal{F}} + \frac{\partial J}{\partial \beta_2} \frac{\partial \beta_2}{\partial \mathcal{F}}$ . This implies that  $\left. \frac{\partial J}{\partial \mathcal{F}} \right|_{\mathcal{F}=0} = \left. \frac{\partial J}{\partial \beta_1} \right|_{\beta_1=\beta_2=\beta}$  is intrinsic and we can write (5.49) as

$$\left. \frac{\partial J}{\partial \mathcal{F}} \right|_{\mathcal{F}=0} = \left. \frac{1}{2} \frac{\partial^2 \alpha_2}{\partial \lambda^2} \right|_{(\lambda=0; \beta_1=\beta_2=\beta)}. \quad (5.51)$$

By using the relation (5.33) for  $p = 2$  between  $\alpha_2(\lambda)$  and the infinite-time cumulant per unit time together with the fact that, when  $\beta_1 = \beta_2$ , the stationary state is the equilibrium state, we get the Green-Kubo relation (5.22). We notice that the present derivation of the Green-Kubo relation is very similar to the argument developed by Lebowitz and Spohn in Ref.[1] in the case where there are several independent currents corresponding to several parameters which drive the system out of equilibrium.

### 5.3 Far from equilibrium relations for infinite-time heat cumulants per unit time

In the present subsection we settle generic results ; thus we replace the cumulated heat  $\mathcal{Q}_2(t)$  by a generic cumulated quantity  $X_t$ . According to the Green-Kubo relation (5.22) where, by virtue of (5.23),  $\partial J / \partial \mathcal{F} |_{\mathcal{F}=0}$  can be replaced by  $\lim_{\mathcal{F} \rightarrow 0} J / \mathcal{F}$ , in the vicinity of equilibrium the ratio  $J / \mathcal{F}$  is equal to the equilibrium value of the second cumulant of the cumulated quantity  $X_t$  per unit time. On the contrary, when the system is far from equilibrium,  $\partial J / \partial \mathcal{F}$  and  $J / \mathcal{F}$  are expected not to be equal to 1/2 times the second cumulant of the cumulated heat  $X_t$  per unit time in the stationary state. The property

$$\frac{\partial J}{\partial \mathcal{F}} \neq \frac{1}{2} \lim_{t \rightarrow +\infty} \frac{\langle X_t^2 \rangle_{st} - \langle X_t \rangle_{st}^2}{t} \quad (5.52)$$

arises from (5.20), while the discrepancy

$$\frac{J}{\mathcal{F}} \neq \frac{1}{2} \lim_{t \rightarrow +\infty} \frac{\langle X_t^2 \rangle_{st} - \langle X_t \rangle_{st}^2}{t} \quad (5.53)$$

is already mentioned in Ref.[1]. This discrepancy implies that the generating function  $\alpha(\lambda)$  for the infinite-time cumulants of  $X_t$  per unit time is not a quadratic function of its argument  $\lambda$ , namely the large deviation function of the current  $\mathcal{J}_t = X_t/t$  is not a quadratic in the generic case, i.e. the probability distribution of  $X_t$  is not asymptotically Gaussian. Indeed, if the generating function  $\alpha(\lambda)$  were quadratic in  $\lambda$ , i.e.  $\alpha(\lambda) = \alpha^{(1)}\lambda + \frac{1}{2}\alpha^{(2)}\lambda^2$ , then the symmetry (5.47) would lead to  $\alpha^{(1)} = \frac{1}{2}\alpha^{(2)}\mathcal{F}$ . In the present section we derive an equation hierarchy for the infinite-time cumulants per unit time far from equilibrium, which in particular gives how  $J/\mathcal{F}$  is related to all even cumulants per unit time in the infinite-time limit.

#### 5.3.1 Generalized Green-Kubo relations for cumulants per unit time far from equilibrium

The starting point is (5.46) which we rewrite for convenience in the generic form

$$\alpha(\lambda; \mathcal{F}) = \alpha(-\mathcal{F} - \lambda; \mathcal{F}), \quad (5.54)$$

where  $\mathcal{F}$  is the thermodynamical force (an inverse temperature difference in (5.43)). The viewpoint is that the coefficients in the expansion of  $\alpha(\lambda; \mathcal{F})$  in powers of  $\lambda$  at  $\lambda = 0$  are related to the heat cumulants per unit time in the infinite-time limit,

$$\bar{\kappa}^{[p]} \equiv \lim_{t \rightarrow +\infty} \frac{1}{t} \kappa^{[p]}(t), \quad (5.55)$$

and these cumulants per unit time are experimentally measurable. This is true at least for the first few ones.

There are a number of ways to rewrite the symmetry (5.54). Formally it is equivalent to

$$e^{-\mathcal{F} \frac{\partial}{\partial \lambda}} \alpha(\lambda; \mathcal{F}) = \alpha(-\lambda; \mathcal{F}). \quad (5.56)$$

Expanding both sides in powers of  $\lambda$  and comparing we get that the cumulants per unit time, given by  $\alpha(\lambda; \mathcal{F}) \equiv \sum_{p=0}^{+\infty} \frac{1}{p!} \lambda^p \bar{\kappa}^{[p]}(\mathcal{F})$  obey the equation hierarchy

$$\bar{\kappa}^{[p]}(\mathcal{F}) = \sum_{q=0}^{+\infty} (-1)^{p+q} \frac{\mathcal{F}^q}{q!} \bar{\kappa}^{[p+q]}(\mathcal{F}) \quad \text{for } p = 0, 1, \dots. \quad (5.57)$$

In (5.57)  $\bar{\kappa}^{[0]}(\mathcal{F}) = 0$ , because  $\alpha(\lambda; \mathcal{F})$  is a cumulant generating function. The relation (5.57) can be explicitly checked in the case of the solvable model studied in paper II [45] in a kinetic regime where the probability distribution of the heat received from the slow thermostat is that of an asymmetric random walk.

A further expansion of the equations (5.57) in powers of  $\mathcal{F}$ , with  $\bar{\kappa}^{[p]}(\mathcal{F}) \equiv \sum_{q=0}^{+\infty} \frac{1}{q!} \mathcal{F}^q \bar{\kappa}^{[p;q]}$ , yields another hierarchy of relations, each of which involves only a finite number of derivatives of cumulants per unit time with respect to  $\mathcal{F}$ ,

$$\bar{\kappa}^{[p;q]} = \sum_{r=0}^q (-1)^{p+r} \binom{q}{r} \bar{\kappa}^{[p+r; q-r]} \quad \text{for } p, q = 0, 1, \dots, \quad (5.58)$$

where  $\binom{q}{r} \equiv q!/[r!(q-r)!]$ , and  $\bar{\kappa}^{[p;0]} = \bar{\kappa}^{[p]}(\mathcal{F} = 0)$  while for  $q \geq 1$

$$\bar{\kappa}^{[p;q]} \equiv \left. \frac{\partial^q \bar{\kappa}^{[p]}}{\partial \mathcal{F}^q} \right|_{\mathcal{F}=0}. \quad (5.59)$$

From the cumulant generating function interpretation, in (5.58) it is to be understood that  $\bar{\kappa}^{[0;q]} = 0$  for  $q = 0, 1, \dots$ , since  $\bar{\kappa}^{[0]}(\mathcal{F}) = 0$ .

The relation hierarchies (5.57) and (5.58) are pure consequences of the symmetry (5.54) without using any further properties of  $\alpha$ . Anyway, they express nothing but a parity relation around  $\lambda = \mathcal{F}/2$ , which fixes half of the coefficients in terms of the other ones, so the equations are not expected to be all independent. Indeed, this is particularly transparent at small order in  $\mathcal{F}$ : for  $q = 0$  (5.58) yields nothing if  $p$  is even, but if  $p$  is odd it leads to  $\bar{\kappa}^{[2n+1;0]} = 0$ . As for the hierarchy (5.57) obtained before expansions in powers of  $\mathcal{F}$ , the dependence between its equations can be exhibited as follows. By specifying the equations to  $p = 2n$  or  $p = 2n + 1$ , the infinite system of equations (5.57) can be rewritten as the equivalent hierarchy,

$$\left. \begin{aligned} \bar{\kappa}^{[2n+1]}(\mathcal{F}) &= \sum_{q=1}^{+\infty} (-1)^{q+1} \frac{\mathcal{F}^q}{(q+1)!} \bar{\kappa}^{[2n+q+1]}(\mathcal{F}) \\ \bar{\kappa}^{[2n+1]}(\mathcal{F}) &= \frac{1}{2} \sum_{q=1}^{+\infty} (-1)^{q+1} \frac{\mathcal{F}^q}{q!} \bar{\kappa}^{[2n+q+1]}(\mathcal{F}) \end{aligned} \right\} \quad \text{for } n = 0, 1, \dots, \quad (5.60)$$

where the right-hand sides of both relations involves both odd and even cumulants per unit time. The difference between the two expressions for  $\bar{\kappa}^{[2n+1]}(\mathcal{F})$  gives the infinite sum rule

$$0 = \sum_{q=2}^{+\infty} (-1)^{q+1} \left[ \frac{1}{(q+1)!} - \frac{1}{2} \frac{1}{q!} \right] \mathcal{F}^q \bar{\kappa}^{[2n+q+1]}(\mathcal{F}).$$

It is not difficult to obtain independent relations which express odd cumulants in terms of even cumulants. We rewrite (5.56) as

$$\left(1 + e^{-\mathcal{F} \frac{\partial}{\partial \lambda}}\right) \alpha(\lambda; \mathcal{F}) = \alpha(\lambda; \mathcal{F}) + \alpha(-\lambda; \mathcal{F}), \quad (5.61)$$

i.e.

$$\alpha(\lambda; \mathcal{F}) = \frac{1}{1 + e^{-\mathcal{F} \frac{\partial}{\partial \lambda}}} [\alpha(\lambda; \mathcal{F}) + \alpha(-\lambda; \mathcal{F})]. \quad (5.62)$$

Note that  $[1 + e^{-x}]^{-1} + [1 + e^x]^{-1} = 1$  so that  $[1 + e^{-x}]^{-1} - \frac{1}{2}$  is an odd function of  $x$ . Hence the Taylor expansion in  $x$  can be written

$$\frac{1}{1 + e^{-x}} = \frac{1}{2} \left(1 + \sum_{k=0}^{+\infty} d_k x^{2k+1}\right) = \frac{1}{2} \left(1 + \frac{x}{2} - \frac{x^3}{24} + \frac{x^5}{240} - \frac{17x^7}{40320} + \dots\right), \quad (5.63)$$

where the coefficients  $d_k$  are related to the classical Bernoulli numbers by

$$d_k = 2 \frac{4^{k+1} - 1}{(2k+2)!} B_{2k+2}. \quad (5.64)$$

From the identity

$$1 = \frac{1}{2}(1 + e^{-x}) \left(1 + \sum_{k=0}^{+\infty} d_k x^{2k+1}\right) \quad (5.65)$$

one infers the recursion relation

$$d_k = \frac{1}{2} \left( \frac{1}{(2k+1)!} - \sum_{0 \leq l < k} \frac{d_l}{(2(k-l))!} \right). \quad (5.66)$$

Finally, in the expansion of (5.62) the even terms cancel out and one is left with

$$\bar{\kappa}^{[2n+1]}(\mathcal{F}) = \sum_{k=0}^{+\infty} d_k \mathcal{F}^{2k+1} \bar{\kappa}^{[2(n+k+1)]}(\mathcal{F}) \quad \text{for } n = 0, 1, \dots, \quad (5.67)$$

which expresses systematically odd cumulants per unit time in terms of even cumulants per unit time. For instance, in the case  $n = 0$ , the ratio of the out-of-equilibrium current  $J(\mathcal{F}) = \bar{\kappa}^{[1]}$  and the thermodynamic force  $\mathcal{F}$  is determined by all even cumulants as

$$\frac{J(\mathcal{F})}{\mathcal{F}} = \sum_{q=0}^{+\infty} d_q \mathcal{F}^{2q} \lim_{t \rightarrow +\infty} \frac{\kappa^{[2(q+1)]}(\mathcal{F})}{t}. \quad (5.68)$$

The latter relation may be viewed as a far-from-equilibrium generalization of the Green-Kubo relation (5.22)-(5.23).

### 5.3.2 Relations between non-linear responses of cumulants per unit time near equilibrium

It is to be noted that the two equivalent hierarchies (5.57) and (5.67) for relations between cumulants make sense only if the cumulants satisfy some growth condition, the second one being most stringent (note that  $d_q \simeq 4(-1)^q \pi^{-2q-2}$ ). Such problem do not arise in the double expansions in both  $\lambda$  and  $\mathcal{F}$ . Expanding (5.67) in powers of  $\mathcal{F}$  yields for  $n = 0, 1, \dots$

$$\begin{aligned} \bar{\kappa}^{[2n+1;0]} &= 0 \\ \bar{\kappa}^{[2n+1;q]} &= \sum_{k=0}^{\lfloor (q-1)/2 \rfloor} \frac{d_k q!}{[q - (2k+1)]!} \bar{\kappa}^{[2(n+k+1);q-(2k+1)]} \quad q = 1, 2, \dots \end{aligned} \quad (5.69)$$

where  $[y]$  denotes the lower integer part of  $y$ . It is easy to check, at least for small values of  $q$ , that the contents of (5.58) and (5.69) are the same. The physical consequences of the latter relations are more readily inferred by explicitly rewriting the relations in the case where  $q$  is odd or even and in terms of either equilibrium cumulants per unit time or the partial derivatives of out-of-equilibrium cumulants per unit time with respect to the thermodynamic force  $\mathcal{F}$ . Then the formulae (5.69) read

$$\bar{\kappa}_{\text{eq}}^{[2n+1]} = 0 \quad (5.70a)$$

$$\left. \frac{\partial \bar{\kappa}^{[2n+1]}}{\partial \mathcal{F}} \right|_{\mathcal{F}=0} = \frac{1}{2} \bar{\kappa}_{\text{eq}}^{[2(n+1)]} \quad (5.70b)$$

$$\left. \frac{\partial^{2m} \bar{\kappa}^{[2n+1]}}{\partial \mathcal{F}^{2m}} \right|_{\mathcal{F}=0} = \sum_{m \geq 1}^m d_{m-r} \frac{(2m)!}{(2r-1)!} \left. \frac{\partial^{2r-1} \bar{\kappa}^{[2(n+m+1-r)]}}{\partial \mathcal{F}^{2r-1}} \right|_{\mathcal{F}=0} \quad (5.70c)$$

$$\left. \frac{\partial^{2m+1} \bar{\kappa}^{[2n+1]}}{\partial \mathcal{F}^{2m+1}} \right|_{\mathcal{F}=0} = \sum_{m \geq 1} d_m (2m+1)! \bar{\kappa}_{\text{eq}}^{[2(n+m+1)]} + \sum_{r=1}^m d_{m-r} \frac{(2m+1)!}{(2r)!} \left. \frac{\partial^{2r} \bar{\kappa}^{[2(n+m+1-r)]}}{\partial \mathcal{F}^{2r}} \right|_{\mathcal{F}=0} \quad (5.70d)$$

where the values of the  $d_k$ 's are given in (5.64).

The equilibrium statements (5.70a) are exemplified in the case of thermal contact as follows. The first cumulant per unit time  $\bar{\kappa}^{[1]}$  for the cumulated heat  $\mathcal{Q}_2$  coincides with the average of the instantaneous current of  $j_2$ ,  $\langle j_2 \rangle_{\text{st}}^{(\beta_1, \beta_2)} \equiv J$ , and from the first equation (5.70a) for  $n = 0$  we retrieve that at equilibrium the mean instantaneous current vanishes,  $J_{\text{eq}} = 0$ . More generally, (5.70a) states that at equilibrium all odd cumulants per unit time of the heat amounts received from one of the two heat baths vanish.

The relations (5.70b) are generalized fluctuation-dissipation relations in the vicinity of equilibrium, which are called “generalized” in the sense that they express the linear response of any cumulant per unit time. Indeed, in the case of thermal contact, from (5.70b) for  $n = 0$  one retrieves the fluctuation-dissipation relation (also called Green-Kubo relation) in the form (5.22) where  $X_t$  is the heat amount received from one of the two heat baths. More generally, from (5.70b) for any value of  $n$  one gets the generalized fluctuation-dissipation relations

$$\lim_{(\beta_1, \beta_2) \rightarrow (\beta, \beta)} \frac{1}{\beta_1 - \beta_2} \left( \lim_{t \rightarrow +\infty} \frac{\kappa^{[2n+1]}}{t} \right) = \frac{1}{2} \lim_{t \rightarrow +\infty} \frac{\kappa_{\text{eq}}^{[2n+2]}}{t}. \quad (5.71)$$

Analogous relations have been derived in the case of several independent out-of-equilibrium steady currents by Andrieux and Gaspard [51]. We notice that for  $n = 1$ , (5.71) means that, in the limit of vanishing  $\beta_1 - \beta_2$ , the ratio between the out-of-equilibrium third centered moment (non-normalized skewness)  $\langle (\mathcal{Q}_2 - \langle \mathcal{Q}_2 \rangle)^3 \rangle$  per unit time and the thermodynamic force  $\beta_1 - \beta_2$  tends to half the equilibrium fourth cumulant (kurtosis multiplied by the square of the variance)  $\langle (\mathcal{Q}_2 - \langle \mathcal{Q}_2 \rangle_{\text{eq}})^4 \rangle_{\text{eq}} - 3 \langle (\mathcal{Q}_2 - \langle \mathcal{Q}_2 \rangle_{\text{eq}})^2 \rangle_{\text{eq}}$  per unit time. These cumulants are expected to be experimentally measurable.

The relations (5.70c) and (5.70d) deal with higher-order response coefficients. The relation (5.70c) in the case  $m = 1$  reads  $\partial^2 \bar{\kappa}^{[2n+1]} / \partial \mathcal{F}^2|_{\mathcal{F}=0} = \partial \bar{\kappa}^{[2n+2]} / \partial \mathcal{F}|_{\mathcal{F}=0}$ . For instance, for  $n = 0$  and  $n = 1$  it leads respectively to

$$\left. \frac{\partial^2 J}{\partial \mathcal{F}^2} \right|_{\mathcal{F}=0} = \left. \frac{\partial \bar{\kappa}^{[2]}}{\partial \mathcal{F}} \right|_{\mathcal{F}=0}, \quad (5.72)$$

and

$$\left. \frac{\partial^2 \bar{\kappa}^{[3]}}{\partial \mathcal{F}^2} \right|_{\mathcal{F}=0} = \left. \frac{\partial \bar{\kappa}^{[4]}}{\partial \mathcal{F}} \right|_{\mathcal{F}=0}. \quad (5.73)$$

The relation (5.70d) in the case  $m = 1$  gives  $\partial^3 \bar{\kappa}^{[2n+1]} / \partial \mathcal{F}^3|_{\mathcal{F}=0} = \frac{3}{2} \partial^2 \bar{\kappa}^{[2n+2]} / \partial \mathcal{F}^2|_{\mathcal{F}=0} - \frac{1}{4} \bar{\kappa}_{\text{eq}}^{[2n+4]}$ . For instance, in the case  $n = 0$  it reads

$$\left. \frac{\partial^3 J}{\partial \mathcal{F}^3} \right|_{\mathcal{F}=0} = \frac{3}{2} \left. \frac{\partial^2 \bar{\kappa}^{[2]}}{\partial \mathcal{F}^2} \right|_{\mathcal{F}=0} - \frac{1}{4} \bar{\kappa}_{\text{eq}}^{[4]}. \quad (5.74)$$



## 6 Extension of previous results to stationary states with several independent currents

In this section we consider the generic case where a system exchanges several kinds of microscopic conserved quantities with several reservoirs and each reservoir exchanges only one kind of these extensive quantities. Moreover, if several reservoirs are involved in a given transition, they should be reservoirs for distinct, independent, physical quantities (this condition is required in the derivation of the MDB.) In the following  $s$  is the index of a species of conserved exchanged quantity (energy, volume, or number of particles ...), while  $a$  is the index of a reservoir. In this situation the microscopic exchange entropy variation  $\delta_{\text{exch}}S(\mathcal{C}' \leftarrow \mathcal{C})$  when the system jumps from configuration  $\mathcal{C}$  to configuration  $\mathcal{C}'$  reads

$$\delta_{\text{exch}}S(\mathcal{C}' \leftarrow \mathcal{C}) = \sum_a F_a \delta x_a(\mathcal{C}' \leftarrow \mathcal{C}). \quad (6.1)$$

The sum runs over all reservoir indices,  $F_a$  is the value of the intensive parameter which characterizes the reservoir with index  $a$  which is at thermodynamic equilibrium, and  $\delta x_a(\mathcal{C}' \leftarrow \mathcal{C})$  is the extensive quantity received by the system from the reservoir  $a$  when the system jumps from  $\mathcal{C}$  to  $\mathcal{C}'$ , with the same convention as in definition (2.20).

### 6.1 Consequences of MDB at finite time

In the form (1.3) where it involves the microscopic exchange entropy variation  $\delta_{\text{exch}}S(\mathcal{C}' \leftarrow \mathcal{C})$  associated with a jump from configuration  $\mathcal{C}$  to configuration  $\mathcal{C}'$  the MDB entails the symmetry (4.6) between the probabilities for time-reversed histories. At a more mesoscopic level, let us compare the probability of evolutions from the configuration  $\mathcal{C}_0$  to the configuration  $\mathcal{C}_f$ , in the course of which the system receives a cumulated quantity  $X_a = \sum \delta x_a$  from each reservoir  $a$ , and the probability of the reversed evolutions, namely evolutions from  $\mathcal{C}_f$  to  $\mathcal{C}_0$  where the cumulated quantities are  $-X_a$ 's. The symmetry (4.8) written in the case of thermal contact becomes in the generic case

$$\frac{P(\mathcal{C}_f | \{X_a\}; t | \mathcal{C}_0)}{P(\mathcal{C}_0 | \{-X_a\}; t | \mathcal{C}_f)} = e^{-\Delta_{\text{exch}}S(\{X_a\})}, \quad (6.2)$$

with

$$\Delta_{\text{exch}}S(\{X_a\}) = \sum_a F_a X_a. \quad (6.3)$$

At the macroscopic level, namely when only the exchanges of extensive quantities with the reservoirs are measured, there appears a symmetry for transient regimes where the system initially prepared in some equilibrium state is suddenly put into contact with reservoirs with thermodynamic parameters  $F_a$ 's that drive the system into a non-equilibrium state. The symmetry involves the excess exchange entropy variation  $\Delta_{\text{exch}}^{\text{excs}}S$ , defined as the difference between the exchange entropy variation under the non-equilibrium external constraints  $F_a$ 's, namely  $\sum_a F_a X_a$ , and the corresponding variation under the equilibrium conditions where for all reservoirs which exchange the same species  $s$  the thermodynamic parameter has the same value  $F_0^{(s)}$ ,

$$\Delta_{\text{exch}}^{\text{excs}}(\{X_a\}; \{F_0^{(s)}\}) \equiv \Delta_{\text{exch}}S(\{X_a\}) - \sum_s F_0^{(s)} \sum_{a \in R(s)} X_a, \quad (6.4)$$

where  $R(s)$  is the set of reservoirs which exchange the microscopic quantity of species  $s$ . When the system is prepared in the equilibrium state with probability distribution  $P_{\text{eq}}(\{F_0^{(s)}\})$  and put into contact with reservoirs with thermodynamic parameters  $F_a$ 's at the initial time of the measurements of exchanged quantities, the excess exchange entropy variation  $\Delta_{\text{exch}}^{\text{excs}}S$  obeys the symmetry relation at any finite time, or “detailed fluctuation relation”,

$$\frac{P_{P_{\text{eq}}(\{F_0^{(s)}\})}(\Delta_{\text{exch}}^{\text{excs}}S)}{P_{P_{\text{eq}}(\{F_0^{(s)}\})}(-\Delta_{\text{exch}}^{\text{excs}}S)} = e^{-\Delta_{\text{exch}}^{\text{excs}}S}. \quad (6.5)$$

The latter relation itself entails the identity, or “integral fluctuation relation”,

$$\langle e^{\Delta_{\text{exch}}^{\text{excs}} S} \rangle_{P_{\text{eq}}(\{F_0^{(s)}\})} = 1. \quad (6.6)$$

## 6.2 Consequences of MDB in the infinite-time limit

A generalization of the argument in subsection 5.1.1 shows that the symmetry relation (6.2) enforced by the MDB at finite time leads to the existence of lower and upper bounds for the ratio between the finite-time probability to measure cumulated quantities with values  $\{X_a\}$  and the corresponding probability to measure the opposite values  $\{-X_a\}$ , when the system is in its stationary state. The finite-time inequalities (5.1) become in the generic case

$$\frac{P_{\text{st}}^{\text{min}}}{P_{\text{st}}^{\text{max}}} \leq \frac{P_{\text{st}}(\{X_a\}; t)}{P_{\text{st}}(\{-X_a\}; t) e^{-\Delta_{\text{exch}} S(\{X_a\})}} \leq \frac{P_{\text{st}}^{\text{max}}}{P_{\text{st}}^{\text{min}}}. \quad (6.7)$$

As a consequence, the dimensionless exchange entropy variation  $\Delta_{\text{exch}} S$  given by (6.3) obeys the fluctuation relation (1.8).

As for the characteristic function of the extensive exchanged quantities  $X_a$ ’s in the stationary state, namely  $\langle e^{\sum_a \lambda_a X_a(t)} \rangle_{\text{st}}$ , the symmetry (6.2)-(6.3) arising from the MDB entails that the characteristic function obeys an inequality similar to (5.41). As a consequence, the generating function of the infinite-time limit of the joint cumulants per unit time obeys a symmetry which is a generalization of (5.42)

$$\alpha_{\{X_a\}}(\{\lambda_a\}) = \alpha_{\{X_a\}}(\{-F_a - \lambda_a\}). \quad (6.8)$$

However, because of the conservation laws for the microscopic quantities exchanged with the reservoirs, the joint cumulants of the  $X_a$ ’s per unit time are not independent in the infinite-time limit. For instance, if  $R(s)$  denotes the set of reservoirs which exchange the microscopic quantity of species  $s$ , because of the conservation law for every species of exchanged quantity,  $\sum_{a \in R(s)} X_a$  is equal to the difference between the values of some observable of the system in the final and initial configurations of the evolution. When the system has a finite number of configurations,  $\sum_{a \in R(s)} X_a$  is bounded and the mean currents  $J_a$ ’s defined as  $J_a = \lim_{t \rightarrow +\infty} \langle X_a(t) \rangle / t$  are related by

$$\sum_{a \in R(s)} J_a = 0. \quad (6.9)$$

Other relations may arise from other microscopic conservation rules determined by the specific forms of the transitions rates. We consider the generic (but not universal) case where there exists a set of cumulated quantities  $Y_\gamma$ ’s, which are linear combinations of the  $X_a$ ’s, but less numerous than the  $X_a$ ’s, and a set of parameters  $\mathcal{F}_\gamma$ ’s such that  $\sum_a F_a X_a - \sum_\gamma \mathcal{F}_\gamma Y_\gamma$  is bounded by a constant independent of time. With these assumptions the expression for the exchange entropy flow in the stationary state takes the generic form

$$\left. \frac{d_{\text{exch}} S}{dt} \right|_{\text{st}} = - \sum_a F_a J_a = - \sum_\gamma \mathcal{F}_\gamma J_\gamma^*, \quad (6.10)$$

where the mean currents  $J_\gamma^*$ ’s, defined as  $J_\gamma^* \equiv \lim_{t \rightarrow +\infty} \langle Y_\gamma(t) \rangle / t$ , are less numerous than the  $J_a$ ’s. Moreover, these assumptions entail that the symmetry (6.2)-(6.3) leads to the inequality

$$m \leq \frac{P(\mathcal{C}_f | \{Y_\gamma\}; t | \mathcal{C}_0)}{P(\mathcal{C}_0 | \{-Y_\gamma\}; t | \mathcal{C}_f) e^{-\sum_\gamma \mathcal{F}_\gamma Y_\gamma}} \leq M, \quad (6.11)$$

where  $m$  and  $M$  are constants. As a consequence, the characteristic function for the  $Y_\gamma$ ’s obeys an inequality similar to (5.41) and subsequently the generating function of the infinite-time limits of their joint cumulants per unit time obeys a symmetry which is a generalization of (5.42),

$$\alpha_{\{Y_\gamma\}}(\{\lambda_\gamma\}; \{\mathcal{F}_\gamma\}) = \alpha_{\{Y_\gamma\}}(\{-\mathcal{F}_\gamma - \lambda_\gamma\}; \{\mathcal{F}_\gamma\}). \quad (6.12)$$

In the framework of graph theory (see subsection 3.3.5) Andrieux and Gaspard [43] have shown an analogous symmetry for dimensionless cumulated quantities  $Y_\gamma$ 's, where each  $Y_\gamma$  is defined as the sum of the cumulated currents through the chords of all cycles in the graph which have the same affinity  $A_\gamma$  and where the  $\mathcal{F}_\gamma$ 's are replaced by the cycle affinities  $A_\gamma$ 's. A similar result is also obtained in Ref.[72]. The latter symmetries are valid even when several reservoirs which exchange the same species can make the system jump from one configuration to another one, as it is the case for coupled chemical reactions.

### 6.3 Generalized Green-Kubo relations for several independent currents

In standard fluctuation-dissipation Green-Kubo formulae, valid near equilibrium, the relevant out-of-equilibrium quantity which is related to equilibrium fluctuations is the Onsager coefficient  $L_{\alpha\gamma}$  introduced in phenomenological thermodynamics of irreversible phenomena as the derivative of a “thermodynamic flux”  $J_\gamma^*$  with respect to a “thermodynamic force”  $\mathcal{F}_\gamma$ , which are associated one to another from the expression of the entropy production rate  $d_{\text{int}}S/dt|_{\text{st}} = -d_{\text{exch}}S/dt|_{\text{st}}$  given by (6.10) when the  $J_\gamma^*$ 's are independent currents between reservoirs. The Onsager coefficient is defined as

$$L_{\alpha\gamma} \equiv \left. \frac{\partial J_\alpha^*}{\partial \mathcal{F}_\gamma} \right|_{\{\mathcal{F}_{\gamma'}=0\}}, \quad (6.13)$$

and the generic statement of Green-Kubo relations reads

$$L_{\alpha\gamma} = \frac{1}{2} \lim_{t \rightarrow +\infty} \frac{\langle Y_\alpha(t) Y_\gamma(t) \rangle_{\text{eq}} - \langle Y_\alpha(t) \rangle_{\text{eq}} \langle Y_\gamma(t) \rangle_{\text{eq}}}{t}. \quad (6.14)$$

Therefore in the case where the non equilibrium stationary state involves several independent stationary currents, the symmetry  $\langle Y_\alpha(t) Y_\gamma(t) \rangle_{\text{eq}} = \langle Y_\gamma(t) Y_\alpha(t) \rangle_{\text{eq}}$  allows to retrieve the phenomenological Onsager symmetry for the off-diagonal Onsager coefficients, namely  $L_{\alpha\gamma} = L_{\gamma\alpha}$ .

Far from equilibrium generalized Green-Kubo relations can be derived from the symmetry (6.12) of the generating function for the infinite-time cumulants per unit time. The generalization of the combinatorics considerations of subsection 5.3 is straightforward. Suppose the independent currents are indexed by  $\gamma \in \Gamma$  so that instead of a single parameter  $\lambda$  one deals with a collection  $\underline{\lambda} \equiv (\lambda_\gamma)_{\gamma \in \Gamma}$ . In the same way, one has a collection  $\underline{\mathcal{F}} \equiv (\mathcal{F}_\gamma)_{\gamma \in \Gamma}$ . However, one considers a single function  $\alpha$  and we rewrite (6.12) as

$$\alpha(\underline{\lambda}; \underline{\mathcal{F}}) = \alpha(-\underline{\mathcal{F}} - \underline{\lambda}; \underline{\mathcal{F}}). \quad (6.15)$$

One could redo all the derivations performed in subsection 5.3. We content to express all odd cumulants in terms of the even ones. Write the expansion in powers of the  $\lambda_\gamma$ 's in a compact way as

$$\alpha(\underline{\lambda}; \underline{\mathcal{F}}) \equiv \sum_{\underline{p} \geq \underline{0}} \frac{1}{\underline{p}!} \underline{\lambda}^{\underline{p}} \bar{\kappa}[\underline{p}](\underline{\mathcal{F}}), \quad (6.16)$$

where the summation is over  $\Gamma$ -tuples  $\underline{p} \equiv (p_\gamma)_{\gamma \in \Gamma}$  of non-negative integers,  $\underline{p}! \equiv \prod_{\gamma \in \Gamma} p_\gamma!$  and  $\underline{\lambda}^{\underline{p}} \equiv \prod_{\gamma \in \Gamma} \lambda_\gamma^{p_\gamma}$ . Note that

$$\bar{\kappa}[\underline{p}](\underline{\mathcal{F}}) \equiv \left. \frac{\partial^{|\underline{p}|} \alpha(\underline{\lambda}; \underline{\mathcal{F}})}{\partial \underline{\lambda}^{\underline{p}}} \right|_{\underline{\lambda}=\underline{0}}, \quad (6.17)$$

where  $|\underline{p}| \equiv \sum_{\gamma \in \Gamma} p_\gamma$  and  $\partial \underline{\lambda}^{\underline{p}} \equiv \prod_{\gamma \in \Gamma} \partial \lambda_\gamma^{p_\gamma}$ . We expand the symmetry relation written in the form

$$\alpha(\underline{\lambda}; \underline{\mathcal{F}}) = \frac{1}{1 + e^{-\underline{\mathcal{F}} \frac{\partial}{\partial \underline{\lambda}}}} [\alpha(\underline{\lambda}; \underline{\mathcal{F}}) + \alpha(-\underline{\lambda}; \underline{\mathcal{F}})], \quad (6.18)$$

where

$$\underline{\mathcal{F}} \frac{\partial}{\partial \underline{\lambda}} \equiv \sum_{\gamma \in \Gamma} \mathcal{F}_\gamma \frac{\partial}{\partial \lambda_\gamma}. \quad (6.19)$$

From

$$\frac{1}{1 + e^{-\mathcal{F}\frac{\partial}{\partial \Delta}}} = \frac{1}{2} \left( 1 + \sum_{k=0}^{+\infty} d_k \sum_{\underline{p} \geq \underline{0}, |\underline{p}|=2k+1} \frac{(2k+1)!}{\underline{p}!} \mathcal{F}^{\underline{p}} \frac{\partial^{2k+1}}{\partial \Delta^{\underline{p}}} \right) \quad (6.20)$$

one infers that, for  $\underline{n}$  such that  $|\underline{n}|$  is odd

$$\bar{\kappa}^{[\underline{n}]}(\mathcal{F}) = \sum_{k=0}^{+\infty} d_k \sum_{\underline{p} \geq \underline{0}, |\underline{p}|=2k+1} \frac{(2k+1)!}{\underline{p}!} \mathcal{F}^{\underline{p}} \bar{\kappa}^{[\underline{n}+\underline{p}]}(\mathcal{F}). \quad (6.21)$$

Again, these relations can be expanded in powers of  $\mathcal{F}$ . The corresponding coefficients of the powers in  $\mathcal{F}$  are the non-linear response coefficients. The relations between the latter non-linear response coefficients are derived by another method in [51]. For instance relations analogous to (5.70d) relate non-linear response coefficients of the cumulants per unit time caused by a variation in thermodynamic forces  $\mathcal{F}_\alpha$ 's to some equilibrium cumulant per unit time, which is symmetric under permutations of the associated cumulated quantities. The latter symmetry is at the origin of Onsager reciprocity relations in the case of  $L_{\alpha\gamma}$  [1] (see definition (6.13)) and of generalized symmetry relations for higher-order response coefficients of cumulants per unit time as noted by Andrieux and Gaspard [73, 51].

## A Property of the coarse grained dynamics

In the present subsection we derive the property (2.5) valid over a period of the ergodic deterministic microscopic dynamics  $\mathcal{T}$  when  $\mathcal{T}$  respects both the conservation of  $E_{\text{dec}}$  and the interaction pattern specified at the end of section 2.1.2.

Though we believe that a general development could be pursued, we prefer to concentrate on a specific model at this point. The system is made of two large parts and a small one, which is reduced to two Ising spins  $\sigma_1, \sigma_2 = \pm 1$ , each one directly in contact with one of the large parts. So a configuration  $C$  can be written as  $C = (C_1, \sigma_1, \sigma_2, C_2)$ . We assume that, when the small part is isolated, its energy  $\mathcal{E}(\sigma_1, \sigma_2)$  does not describe independent spins. Moreover, the microscopic dynamics  $\mathcal{T}$  conserves the energy  $E_{\text{dec}}(C)$  where the interactions between parts is neglected, namely  $E_{\text{dec}}(C) \equiv E_1(C_1) + E_2(C_2) + \mathcal{E}(\sigma_1, \sigma_2)$ . The remnant of interactions between the parts is embodied in the following restrictions. For any  $C$ ,  $\mathcal{T}(C)$  is obtained from  $C$  by one of the following operations :

- (I) a flip of spin  $\sigma_1$  together with a change in  $C_1$  and a possible change in  $C_2$  such that  $E_1(C_1) + \mathcal{E}(\sigma_1, \sigma_2)$  and  $E_2(C_2)$  both remain constant (i.e. the energy needed to flip the spin  $\sigma_1$  entirely comes from or goes to the large part 1).
- (II) a flip of spin  $\sigma_2$  together with a change in  $C_2$  and a possible change in  $C_1$  such that  $E_1(C_1)$  and  $E_2(C_2) + \mathcal{E}(\sigma_1, \sigma_2)$  both remain constant (i.e. the energy needed to flip the spin  $\sigma_2$  entirely comes from or goes to large part 2).
- (III) a change in  $C_1$  and/or  $C_2$  but no flip of  $\sigma_1$  or  $\sigma_2$ , such that  $E_1(C_1)$  and  $E_2(C_2)$  both remain constant, as well as  $\mathcal{E}(\sigma_1, \sigma_2)$ .

In order to build an effective mesoscopic dynamics we just keep track of  $(E_1, \sigma_1, \sigma_2, E_2)$  as a function of time. Therefore during the time evolution of a given configuration  $C$  of the whole system we concentrate only on time steps at which a change of type (I) or (II) occurs, namely when either spin  $\sigma_1$  or spin  $\sigma_2$  is flipped with known corresponding variations in  $E_1$  and  $E_2$ . We do not follow precisely the changes (III) that modifies the configurations of large parts without changing the energy of any part (either  $E_1$ ,  $E_2$  or  $\mathcal{E}(\sigma_1, \sigma_2)$ ). The possible changes are

$$(E_1, \sigma_1, \sigma_2, E_2) \rightarrow (E'_1, -\sigma_1, \sigma_2, E_2) \quad \text{with} \quad E'_1 = E_1 + \mathcal{E}(\sigma_1, \sigma_2) - \mathcal{E}(-\sigma_1, \sigma_2) \quad \text{type (I)} \quad (\text{A.1})$$

and

$$(E_1, \sigma_1, \sigma_2, E_2) \rightarrow (E_1, \sigma_1, -\sigma_2, E'_2) \quad \text{with} \quad E'_2 = E_2 + \mathcal{E}(\sigma_1, \sigma_2) - \mathcal{E}(\sigma_1, -\sigma_2) \quad \text{type (II)} \quad (\text{A.2})$$

Starting from some initial configuration  $(E_1^0, \sigma_1^0, \sigma_2^0, E_2^0)$ , some set of possible  $(E_1, \sigma_1, \sigma_2, E_2)$  will be visited during the time evolution over a period of  $\mathcal{T}$ , and it is useful to view this set as the vertices of a graph, whose edges connect two vertices if one can jump from one to the other by a single change of type (I) or (II). This graph can be chosen to be unoriented because a transformation of type (I) or of type (II) is its own inverse. Then a trajectory over a period of the underlying deterministic dynamics corresponds to a closed walk on this graph, which summarizes the coarse graining due to the macroscopic description of the large parts. During a period of the microscopic dynamics the closed walk on the graph goes through each edge a number of times.

The main observation is that, since energy  $\mathcal{E}(\sigma_1, \sigma_2)$  does not describe independent spins, the graph has the topology of a segment. Indeed, the graph is connected by construction, but each vertex has at most two neighbors. So the graph is either a segment or a circle. Let us suppose that it is a circle. Then one can go from  $(E_1^0, \sigma_1^0, \sigma_2^0, E_2^0)$  to itself by visiting each edge exactly once, i.e. by an alternation of moves of type (I) and (II). After two steps both spins in the small part are flipped, so the total length of the circle is a multiple of 4. But after 4 steps a definite amount of energy has been transferred between the two large parts, namely the energy in the first large part has changed by  $\mathcal{E}(\sigma_1, \sigma_2) - \mathcal{E}(-\sigma_1, \sigma_2) + \mathcal{E}(-\sigma_1, -\sigma_2) - \mathcal{E}(\sigma_1, -\sigma_2)$  and a trivial computation shows that this cannot vanish unless  $\mathcal{E}(\sigma_1, \sigma_2)$  describes two independent spins, a possibility which has been discarded. This excludes the circle topology.

Since the graph is a segment, any closed walk on the graph traverses a given edge the same number of times in one direction and in the other one. As a consequence, during a period of the microscopic dynamics  $\mathcal{T}$ , the motion induced by  $\mathcal{T}$  on the graph is such that the number of transitions of type (I)  $(E_1, \sigma_1, \sigma_2, E_2) \rightarrow (E'_1, -\sigma_1, \sigma_2, E_2)$  is equal to the number of their inverse transitions  $(E'_1, -\sigma_1, \sigma_2, E_2) \rightarrow (E_1, \sigma_1, \sigma_2, E_2)$ , where the relation between  $E'_1$  and  $E_1$  is given in (A.1). The same considerations apply for transition rates associated with the flipping of  $\sigma_2$ . The result is summarized in (2.5).

For a more general discrete system, the mesoscopic time evolution can again be characterized by the time steps where the small system variables are flipped while the changes in the large parts are only macroscopically described by their net energies. An analogous graph can be constructed but it can be much harder to analyze its topology, which is the crucial knowledge needed to exploit the consequences of ergodicity. These consequences are most stringent for a tree. There is no reason a priori why the graph should be a tree, but what this implies, namely that ratios for the transition rates involving the small part and a large part are given by ratios of energy level degeneracies in the large part, is physically quite appealing.

## B The Markovian approximation

This appendix is a short digression on mathematics. The aim is to briefly recall a trick allowing to replace a fixed sequence of symbols by a random one with “analogous” statistical properties, and then to combine this trick with a coarse graining procedure.

### B.1 Discrete time stochastic approximation

Suppose  $w \equiv x_1 x_2 \cdots x_N$  is a given finite sequence of elements of a finite set  $S$ . It is convenient to assume that this sequence is periodic, i.e. that  $x_{N+1} \equiv x_1$ . For  $x \in S$ , set  $N_x \equiv \#\{i \in [1, N], x_i = x\}$ , i.e.  $N_x$  is the number of occurrences of  $x$  in the sequence  $w$ . There is no loss of generality in assuming that  $N_x \neq 0$  for every  $x \in S$ , should it lead to consider a smaller  $S$ . For  $x, x' \in S$  set  $N_{xx'} \equiv \#\{i \in [1, N], x_i = x, x_{i+1} = x'\}$ , i.e.  $N_{xx'}$  is the number of occurrences of the pattern  $xx'$  in the sequence  $w$ . Notice that we accept that  $x = x'$  in this definition. Of course, we could look at the occurrence of more general patterns. By definition, we have  $\sum_{x' \in S} N_{xx'} = \sum_{x' \in S} N_{x'x} = N_x$ . The result we want to recall is the following.

There is a single time-homogeneous irreducible Markov matrix on  $S$  which fulfills the following two requirements. First, in the stationary state of the corresponding discrete-time stochastic evolution, the probability  $P_{\text{st}}(x', i+1; x, i)$  that a sample  $\hat{x}$  takes the value  $x$  at time  $i$  and the

value  $x'$  at time  $i + 1$  is equal to the frequency of the pattern  $xx'$  in the sequence  $w$ , namely

$$P_{\text{st}}(x', i + 1; x, i) = \frac{N_{xx'}}{N}. \quad (\text{B.1})$$

Second, the stationary probability that the random variable  $\hat{x}$  takes the value  $x$  at any time  $i$  is equal to the frequency of  $x$  in the sequence  $w$ , namely

$$P_{\text{st}}(x) = \frac{N_x}{N}. \quad (\text{B.2})$$

In fact (B.2) is a consequence of (B.1), because of the relations  $P_{\text{st}}(x) = \sum_{x' \in S} P_{\text{st}}(x', i + 1; x, i)$  and  $\sum_{x' \in S} N_{xx'} = N_x$ . To say it in words, there is a unique time-homogeneous irreducible Markov chain whose stationary statistics for patterns of length 1 or 2 is the same as the corresponding statistics in  $w$ .

The proof is elementary. By the Markov property, the Markov matrix element from  $x$  to  $x'$ , denoted<sup>4</sup> by  $T(x' \leftarrow x)$ , must satisfy the relation  $P_{\text{st}}(x', i + 1; x, i) = T(x' \leftarrow x)P_{\text{st}}(x)$ , so the only candidate is

$$T(x' \leftarrow x) = \frac{N_{xx'}}{N_x}. \quad (\text{B.3})$$

Conversely, the corresponding matrix is obviously a Markov matrix ( $\sum_{x' \in S} T(x' \leftarrow x) = 1$  since  $\sum_{x' \in S} N_{xx'} = N_x$ ), and it is irreducible, because as  $w$  contains all elements of  $S$ , transitions within  $w$  allow to go from every element of  $S$  to every other. Checking that for this Markov matrix the stationary measure, namely the solution of  $\sum_{x \in S} T(x' \leftarrow x)P_{\text{st}}(x) = P_{\text{st}}(x')$ , is given by (B.2) boils down to the identity  $\sum_{x \in S} N_{xx'} = N_{x'}$  recalled above, and then the two-point property (B.1) follows. This finishes the proof.

In the case very specific case where  $x_1, x_2, \dots, x_N$  are all distinct, i.e.  $|S|$ , the cardinal of  $S$ , is equal to  $N$ ,  $N_x = 1$  for all  $x \in S$  and  $N_{xx'} = \delta_{x', x_{i+1}}$  where  $i$  is such that  $x_i = x$ . Then the only randomness lies in the choice of the initial distribution, and each trajectory of the Markov chain reproduces  $w$  up to a translation of all indices. If  $N$  is large and  $|S|$  is  $\sim N$ , slightly weaker but analogous conclusions survive. A more interesting case is when  $|S| \ll N$  by many orders of magnitude.

This trick has been used for instance to write a random text “the Shakespeare way” by computing the statistics of sequences of two words in one of his books.

By definition, the Markovian approximation reproduces the statistics of the original sequence only for length 1 and length 2 patterns. It is a delicate issue to decide whether or not it also does a reasonable job for other patterns. For instance, the random Shakespeare book certainly looks queer. Various physical but heuristic arguments suggest that for the kind of sequences  $w$  relevant for this work, the Markov approximation is quite good, but we shall not embark on that.

## B.2 Continuous time approximation

By the very same argument, there is a single time-homogeneous irreducible Markov transition matrix on  $S$  and a single time scale  $\tau$  such that, in the stationary state of the corresponding continuous-time stochastic evolution, the expected number of transitions from  $x$  to  $x' \neq x$  per unit time is  $N_{xx'}/\tau$ . The formula one finds for the transition rate is

$$W(x' \leftarrow x) = \frac{N_{xx'}}{\tau N_x} \text{ for } x \neq x'. \quad (\text{B.4})$$

The continuous-time approximation becomes more natural in the case when  $|S|$ , the cardinal of  $S$ , is such that  $|S| \ll N$ , while for all  $x$   $N_{xx} \sim N_x$  and for all  $x \neq x'$   $N_{xx'} \ll N_x$ , which means that transitions are rare and most of the time  $x$  follows  $x$  in the sequence  $w$ . Again, one can argue that for the kind of sequences  $w$  relevant for this work, this is guaranteed by physics. Then taking

---

<sup>4</sup>In the whole paper we use the convention that the evolution (here given by the transition matrix  $T$ ) is written from the right to the left as in quantum mechanics.

$\tau$  (in some macroscopic time unit) of the order of the largest of the  $N_{xx'}/N_x$ ,  $x \neq x'$ , one ends with a Markov transition matrix with elements of order unity (in some macroscopic inverse time unit), and  $t_i = \tau i$  can be taken as the physical macroscopic time.

To summarize, in this work, we shall systematically associate to certain sequences  $w$  a continuous-time Markov process and exploit properties of  $w$  to constraint the structure of  $W(x' \leftarrow x)$ .

A natural application of the above ideas is to the case where the sequence  $w$  arises from some coarse graining procedure. One starts from a sequence  $\omega = \xi_1 \xi_2 \cdots \xi_N$  where  $\xi_1, \xi_2, \cdots, \xi_N$  belong to a set  $\Sigma$  that is so large that  $\omega$  cannot be stored, and that only some of its features can be kept. Say we partition  $\Sigma \equiv \bigcup_{x \in S} \Sigma_x$  where  $S$  is of reasonable size. Then all we keep of  $\omega$  is  $w = x_1 x_2 \cdots x_N$  where, for  $i \in [1, N]$ ,  $x_i$  is substituted for  $\xi_i$  when  $\xi_i \in \Sigma_{x_i}$ . In the applications we have in mind,  $\Sigma$  is an  $N$ -element set, i.e. all terms in the sequence  $\omega$  are distinct. In that case, even if  $\omega$  was constructed in a perfectly deterministic way, by saying who follows who in the sequence, such a description is unavailable on  $w$ , and  $w$  may well look quite random, so the Markov chain approximation is worth a try. In fact,  $|S| \ll N$  and we shall take as a physical input that transitions are rare, so that the (continuous-time) Markov process is an excellent approximation to the (discrete time) Markov chain.

## C Microcanonical detailed balance and time reversal invariance in Hamiltonian dynamics

In view of comparison with the approach developed for discrete variables, we rederive the microcanonical detailed balance (2.9) in the case where a microscopic configuration of the full system is described by continuous variables and the system has a microscopic deterministic dynamics whose Hamiltonian  $H$ , independent of time, is an even function of momenta. (Similar arguments can be found in derivations which rely on different assumptions in Refs.[74, 48].)

Further hypotheses are the following. The coordinate of the system in phase space is denoted by  $\xi$ , and  $\underline{x}[\xi]$  is a set of mesoscopic variables defined from the microscopic coordinate  $\xi$  and which are even functions of momenta. The initial position of the system in phase space is not known, but we assume that it is uniformly randomly distributed in the energy shell  $E = H[\xi]$ , where  $E$  is the value of the energy of the full system at some initial time  $t_0$ . In other words, the initial probability distribution of  $\xi$ ,  $P_0(\xi)$ , is such that

$$d\xi P_0(\xi) \equiv \frac{d\xi \delta(H[\xi] - E)}{\int d\xi' \delta(H[\xi'] - E)}, \quad (\text{C.1})$$

where  $d\xi$  is the Liouville measure in phase space.

The dynamics is invariant by time translation and we simply denote by  $f_t(\xi)$  the position of the system in phase space at time  $t_0 + t$  knowing that it is at position  $\xi$  at time  $t_0$ . (For the same reason, in the following the initial time  $t_0$  is set equal to 0). Then the invariance of the Hamiltonian under the operation  $R$  which changes every momentum into its opposite implies that

$$[R f_t R f_t](\xi) = \xi, \quad (\text{C.2})$$

namely the microscopic dynamics is invariant under time reversal. Since the Hamiltonian is a constant of motion and the equations of motion conserve the infinitesimal volume in phase space, the probability  $d\xi P_0(\xi)$  defined in (C.1) is also conserved under the microscopic evolution,

$$d(f_t[\xi]) P_0(f_t[\xi]) = d\xi P_0(\xi). \quad (\text{C.3})$$

When the initial position of the system in phase space is distributed according to  $P_0$ , the probability that at time  $t$  the set of mesoscopic variable  $\underline{x}$  takes the value  $\underline{x}_1$ , which is defined as

$$P_{P_0}(\underline{x}_1, t) \equiv \int d\xi P_0(\xi) \delta(\underline{x}[f_t(\xi)] - \underline{x}_1), \quad (\text{C.4})$$

is in fact independent of time (i.e. conserved by the microscopic dynamics). Indeed, let us consider the change of variable  $\xi' = f_t(\xi)$ . The conservation of  $P_0(\xi)d\xi$  by the dynamics (C.3) implies that  $d\xi P_0(\xi) = d\xi' P_0(\xi')$ . Then the integral in (C.4) reads  $\int d\xi' P_0(\xi') \delta(\underline{x}[\xi'] - \underline{x}_1)$ . It is independent of time and, by virtue of the definition (C.1) of the distribution  $P_0(\xi)$  in phase space, it is in fact equal to the microcanonical distribution  $P_{\text{mc}}(\underline{x}_1)$ ,

$$P_{P_0}(\underline{x}_1, t) = P_{\text{mc}}(\underline{x}_1) \equiv \frac{\int d\xi \delta(H[\xi] - E) \delta(\underline{x}[\xi] - \underline{x}_1)}{\int d\xi' \delta(H[\xi'] - E)}. \quad (\text{C.5})$$

On the other hand, the probability that, again when the initial position of the system in phase space is distributed according to  $P_0(\xi)$ , the set of mesoscopic variables  $\underline{x}$  takes the value  $\underline{x}_1$  at time  $t = 0$  and the value  $\underline{x}_2$  at time  $t$  reads

$$P_{P_0}(\underline{x}_2, t; \underline{x}_1, 0) = \int d\xi P_0(\xi) \delta(\underline{x}[f_t(\xi)] - \underline{x}_2) \delta(\underline{x}[\xi] - \underline{x}_1). \quad (\text{C.6})$$

Since the variables denoted by  $\underline{x}(\xi)$  are even functions of the momenta, we can write  $\underline{x}[f_t(\xi)] = \underline{x}[[Rf_t](\xi)]$ . Let us consider the change of variable  $\xi' = [Rf_t](\xi)$ . The invariance under time reversal of the microscopic dynamics (C.2) entails that  $\xi = [Rf_t](\xi')$ , so that  $\underline{x}[\xi] = \underline{x}[[Rf_t](\xi')] = \underline{x}[f_t(\xi')]$ . By virtue of its definition (C.1)  $d\xi P_0[\xi]$  is invariant under the operation  $R$ , so  $d\xi' P_0(\xi') = d(f_t[\xi]) P_0(f_t[\xi])$  and the conservation of  $P_0(\xi)d\xi$  by the dynamics (C.3) implies that  $d\xi' P_0(\xi') = d\xi P_0(\xi)$ . Eventually the integral in (C.6) reads  $\int d\xi' P_0(\xi') \delta(\underline{x}[\xi'] - \underline{x}_2) \delta(\underline{x}[f_t(\xi')] - \underline{x}_1)$  and we get

$$P_{P_0}(\underline{x}_2, t; \underline{x}_1, 0) = P_{P_0}(\underline{x}_1, t; \underline{x}_2, 0). \quad (\text{C.7})$$

The joint probability is invariant by exchanging the times at which  $\underline{x}_1$  and  $\underline{x}_2$  occur.

We now assume that the evolution of the probability of the mesoscopic variable  $\underline{x}$ , which is invariant by time translation, can be described by an homogeneous Markovian stochastic process whose stationary distribution  $P_{\text{st}}(\underline{x})$  is the time-independent probability  $P_{P_0}(\underline{x})$  defined in (C.4) and whose transition rates, denoted by  $W(\underline{x}' \leftarrow \underline{x})$ , are determined by the identification

$$P_{P_0}(\underline{x}', dt; \underline{x}, 0) \equiv W(\underline{x}' \leftarrow \underline{x}) P_{P_0}(\underline{x}) \times dt, \quad (\text{C.8})$$

where the time-displaced joint probability for  $\underline{x}'$  and  $\underline{x}$  and the probability of the single variable  $\underline{x}$  are calculated with the same initial distribution  $P_0$  for the microscopic variables  $\xi$ 's. Then, the time reversal symmetry property (C.7) for the joint probability and the fact that  $P_{P_0}(\underline{x})$  coincides with the microcanonical distribution  $P_{\text{mc}}(\underline{x})$  by virtue of (C.5) lead to the microcanonical detailed balance

$$W(\underline{x}' \leftarrow \underline{x}) P_{\text{mc}}(\underline{x}) = W(\underline{x} \leftarrow \underline{x}') P_{\text{mc}}(\underline{x}'). \quad (\text{C.9})$$

## D Definitions for statistics over histories

Consider a history where the system starts in configuration  $\mathcal{C}_0$  at time  $t_0 = 0$  and ends in configuration  $\mathcal{C}_f$  at time  $t$  after going through successive configurations  $\mathcal{C}_0, \mathcal{C}_1, \dots, \mathcal{C}_N = \mathcal{C}_f$ . The  $N$  instantaneous jumps from one configuration to another occur at  $N$  successive intermediate times  $T_i$  which are continuous stochastic variables : the system jumps from  $\mathcal{C}_{i-1}$  to  $\mathcal{C}_i$  at time  $T_i$  ( $i = 1, \dots, N$ ) in the time interval  $[t_i, t_i + dt_i[$ , with  $t_0 < t_1 < t_2 < \dots < t_N < t$ . The probability measure for such a history is related to the probability density  $\Pi_{\mathcal{C}_f, \mathcal{C}_0}[\mathcal{H}ist]$  by

$$dP_{\mathcal{C}_f, \mathcal{C}_0}[\mathcal{H}ist] \equiv dt_1 \dots dt_N \Pi_{\mathcal{C}_f, \mathcal{C}_0}[\mathcal{H}ist] \quad (\text{D.1})$$

where, for a time-translational invariant process,

$$\begin{aligned} \Pi_{\mathcal{C}_f, \mathcal{C}_0}[\mathcal{H}ist] &= e^{-(t-t_N)\Lambda(\mathcal{C}_N)} (\mathcal{C}_N | \mathbb{W} | \mathcal{C}_{N-1}) e^{-(t_N-t_{N-1})\Lambda(\mathcal{C}_{N-1})} \\ &\quad \times \dots e^{-(t_2-t_1)\Lambda(\mathcal{C}_1)} (\mathcal{C}_1 | \mathbb{W} | \mathcal{C}_0) e^{-(t_1-t_0)\Lambda(\mathcal{C}_0)} \end{aligned} \quad (\text{D.2})$$



and  $\Lambda(\mathcal{C})$  is the total exit rate from configuration  $\mathcal{C}$ ,

$$\Lambda(\mathcal{C}) \equiv - \sum_{\mathcal{C}' \neq \mathcal{C}} (\mathcal{C}' | \mathbb{W} | \mathcal{C}). \quad (\text{D.3})$$

The average of a functional  $F[\mathcal{H}ist]$  over the histories which start in configuration  $\mathcal{C}_0$  and end in configuration  $\mathcal{C}_f$  is computed as

$$\langle F \rangle_{\mathcal{C}_f, \mathcal{C}_0} = \int dP_{\mathcal{C}_f, \mathcal{C}_0} [\mathcal{H}ist] F[\mathcal{H}ist] \quad (\text{D.4})$$

with

$$\int dP_{\mathcal{C}_f, \mathcal{C}_0} [\mathcal{H}ist] = \sum_{N=0}^{+\infty} \sum_{\mathcal{C}_1} \dots \sum_{\mathcal{C}_{N-1}} \int_{t_0 < t_1 < \dots < t_N} dP_{\mathcal{C}_f, \mathcal{C}_0} [\mathcal{H}ist]. \quad (\text{D.5})$$

Then the average of a functional when the initial distribution of configurations is  $P_0$  reads

$$\langle F \rangle_{P_0} = \sum_{\mathcal{C}_f} \sum_{\mathcal{C}_0} P_0(\mathcal{C}_0) \int dP_{\mathcal{C}_f, \mathcal{C}_0} [\mathcal{H}ist] F[\mathcal{H}ist]. \quad (\text{D.6})$$

## E Some remarks on large deviations

Suppose that  $X_t$  is some time-dependent random quantity. Typically, what we have in mind is a variation of some physical quantity (heat, entropy) exchanged between a reservoir and a system during the interval  $[0, t]$ . So  $X_t$  is expected to scale like  $t$  at large times, with an average  $\langle X_t \rangle / t$  going to a constant  $J$  as  $t$  goes to  $+\infty$ .

There are a number of definitions to quantify the probability that  $X_t/t$  differs significantly from  $J$  at some large time. It often happens that this probability is exponentially small, and a general mathematical theory, large deviation theory, has emerged to describe this situation.

### E.1 The mathematical definition of large deviations

The general definition is a bit abstract, dealing with general probability measures depending on  $t$ . We restrict here to the situation when the probability measure is the distribution of a real random variable  $X_t$ . If  $j_- < j_+$  are two real numbers, observe that  $P\left(\frac{X_t}{t} \in ]j_-, j_+]\right) \leq P\left(\frac{X_t}{t} \in [j_-, j_+]\right)$  so that

$$-\frac{1}{t} \ln P\left(\frac{X_t}{t} \in [j_-, j_+]\right) \leq -\frac{1}{t} \ln P\left(\frac{X_t}{t} \in ]j_-, j_+]\right). \quad (\text{E.1})$$

For large  $t$ , the two members may not converge but at least

$$\liminf_{t \rightarrow +\infty} -\frac{1}{t} \ln P\left(\frac{X_t}{t} \in [j_-, j_+]\right) \leq \limsup_{t \rightarrow +\infty} -\frac{1}{t} \ln P\left(\frac{X_t}{t} \in ]j_-, j_+]\right), \quad (\text{E.2})$$

where both sides of the inequality belong to  $[0, +\infty]$ . A nontrivial lower bound for the left-hand side or upper bound for the right-hand side gives information on the exponential rate of decrease of  $P\left(\frac{X_t}{t} \in [j_-, j_+]\right)$  and  $P\left(\frac{X_t}{t} \in ]j_-, j_+]\right)$ , the best situation being when the two exist and coincide.

One says that the random variable  $X_t$  obeys a large deviation principle (LDP) if there is a lower semi-continuous function  $R_X$ , called the rate function, or large deviation function, such that

$$\inf_{j \in [j_-, j_+]} R_X(j) \leq \liminf_{t \rightarrow +\infty} -\frac{1}{t} \ln P\left(\frac{X_t}{t} \in [j_-, j_+]\right) \quad (\text{E.3})$$

and

$$\limsup_{t \rightarrow +\infty} -\frac{1}{t} \ln P\left(\frac{X_t}{t} \in ]j_-, j_+]\right) \leq \inf_{j \in ]j_-, j_+]} R_X(j). \quad (\text{E.4})$$

The property that  $R_X$  is lower semicontinuous means that

$$\lim_{\varepsilon \searrow 0} \inf_{j' \in [j-\varepsilon, j+\varepsilon]} R_X(j') = R_X(j) \quad \text{for every } j. \quad (\text{E.5})$$

Notice that using an open interval  $]j-\varepsilon, j+\varepsilon[$  instead of  $[j-\varepsilon, j+\varepsilon]$  in this definition leads to the same notion. By taking  $j_+ = j+\varepsilon$ ,  $j_- = j-\varepsilon$  and letting  $\varepsilon \searrow 0$ , we see that if  $X_t$  obeys a LDP then the rate function is given by

$$\begin{aligned} R_X(j) &= \lim_{\varepsilon \searrow 0} \liminf_{t \rightarrow +\infty} -\frac{1}{t} \ln P \left( \frac{X_t}{t} \in [j-\varepsilon, j+\varepsilon] \right) \\ &= \lim_{\varepsilon \searrow 0} \limsup_{t \rightarrow +\infty} -\frac{1}{t} \ln P \left( \frac{X_t}{t} \in ]j-\varepsilon, j+\varepsilon[ \right). \end{aligned} \quad (\text{E.6})$$

Note that we do not claim that the existence of the above limits, equal to some lower semicontinuous function  $R_X(j)$ , guarantees that  $X_t$  satisfies a LDP. The above definitions involve limsup and liminf and make a difference between  $]j_-, j_+[$  and  $[j_-, j_+]$  to take care of slightly pathological situations.

However, if  $X_t$  satisfies a LDP with a continuous rate function, then  $\inf_{j \in ]j_-, j_+[} R_X(j) = \inf_{j \in [j_-, j_+]} R_X(j)$ , the limsup and liminf become standard limits and the inequalities become equalities :

$$\lim_{t \rightarrow +\infty} -\frac{1}{t} \ln P \left( \frac{X_t}{t} \in ]j_-, j_+[ \right) = \lim_{t \rightarrow +\infty} -\frac{1}{t} \ln P \left( \frac{X_t}{t} \in [j_-, j_+] \right) = \inf_{j \in [j_-, j_+]} R_X(j). \quad (\text{E.7})$$

One can rewrite this informally as

$$P \left( \frac{X_t}{t} \in I \right) \underset{t \rightarrow +\infty}{\sim} e^{-t \inf_{j \in I} R_X(j)} \quad (\text{E.8})$$

whenever  $I$  is an interval. More and more informal, and possibly misleading, statements are

$$P \left( \frac{X_t}{t} \in [j, j+dj] \right) \underset{t \rightarrow +\infty}{\sim} e^{-t R_X(j)} \quad \text{or} \quad P \left( \frac{X_t}{t} = j \right) \underset{t \rightarrow +\infty}{\sim} e^{-t R_X(j)}. \quad (\text{E.9})$$

Before we turn to other possible characterizations of large deviations, let us explain briefly and heuristically the presence of the  $\inf R_X(j)$  in these formulae. If the interval  $I$  is split into a finite number of intervals, say  $I = \cup_{k=1}^K I_k$ , we have

$$\max_k P \left( \frac{X_t}{t} \in I_k \right) \leq P \left( \frac{X_t}{t} \in I \right) \leq \sum_k P \left( \frac{X_t}{t} \in I_k \right) \leq K \max_k P \left( \frac{X_t}{t} \in I_k \right). \quad (\text{E.10})$$

We infer that

$$-\frac{\ln K}{t} + \min_k -\frac{1}{t} \ln P \left( \frac{X_t}{t} \in I_k \right) \leq -\frac{1}{t} \ln P \left( \frac{X_t}{t} \in I \right) \leq \min_k -\frac{1}{t} \ln P \left( \frac{X_t}{t} \in I_k \right) \quad (\text{E.11})$$

so that, for  $s \geq t$ ,

$$-\frac{\ln K}{t} + \min_k -\frac{1}{s} \ln P \left( \frac{X_s}{s} \in I_k \right) \leq -\frac{1}{s} \ln P \left( \frac{X_s}{s} \in I \right) \leq \min_k -\frac{1}{s} \ln P \left( \frac{X_s}{s} \in I_k \right). \quad (\text{E.12})$$

First we take the inf over  $s \geq t$  and interchange freely the inf and the min to get

$$-\frac{\ln K}{t} + \min_k \inf_{s \geq t} -\frac{1}{s} \ln P \left( \frac{X_s}{s} \in I_k \right) \leq \inf_{s \geq t} -\frac{1}{s} \ln P \left( \frac{X_s}{s} \in I \right) \leq \min_k \inf_{s \geq t} -\frac{1}{s} \ln P \left( \frac{X_s}{s} \in I_k \right). \quad (\text{E.13})$$

Letting  $t \rightarrow +\infty$  yields

$$\liminf_{t \rightarrow +\infty} -\frac{1}{t} \ln P \left( \frac{X_t}{t} \in I \right) = \min_k \liminf_{t \rightarrow +\infty} -\frac{1}{t} \ln P \left( \frac{X_t}{t} \in I_k \right). \quad (\text{E.14})$$

Second we take the sup over  $s \geq t$ . This time only one of the inequalities survives the interchange of sup and min

$$\sup_{s \geq t} -\frac{1}{s} \ln P \left( \frac{X_s}{s} \in I \right) \leq \sup_{s \geq t} \min_k -\frac{1}{s} \ln P \left( \frac{X_s}{s} \in I_k \right) \leq \min_k \sup_{s \geq t} -\frac{1}{s} \ln P \left( \frac{X_s}{s} \in I_k \right) \quad (\text{E.15})$$

and letting  $t \rightarrow +\infty$  yields

$$\limsup_{t \rightarrow +\infty} -\frac{1}{t} \ln P \left( \frac{X_t}{t} \in I \right) \leq \min_k \limsup_{t \rightarrow +\infty} -\frac{1}{t} \ln P \left( \frac{X_t}{t} \in I_k \right). \quad (\text{E.16})$$

To summarize

$$\begin{aligned} \min_k \liminf_{t \rightarrow +\infty} -\frac{1}{t} \ln P \left( \frac{X_t}{t} \in I_k \right) &= \liminf_{t \rightarrow +\infty} -\frac{1}{t} \ln P \left( \frac{X_t}{t} \in I \right) \\ \limsup_{t \rightarrow +\infty} -\frac{1}{t} \ln P \left( \frac{X_t}{t} \in I \right) &\leq \min_k \limsup_{t \rightarrow +\infty} -\frac{1}{t} \ln P \left( \frac{X_t}{t} \in I_k \right). \end{aligned} \quad (\text{E.17})$$

Informally, this implies indeed that the large deviation estimates for the interval  $I$  come solely from some arbitrary small subinterval of  $I$  characterized by a minimizing property.

## E.2 Alternative definitions

There are a number of situations of physical interest when  $X_t$  takes values in a (time independent) discrete set  $\mathbb{S}$ . Typically  $\mathbb{S}$  is  $\mathbb{Z}$ , the set of integers, or  $a + \mathbb{Z}$ , the set of integers shifted by  $a$ , or  $a\mathbb{Z}$ , the set of integers dilated by  $a$ .

Though this discreteness by no means prevents from using the general large deviation theory, other reasonable but ad hoc definitions come to mind.

For instance, one may ask whether there is a function  $g_X(j)$  from  $\mathbb{R}$  to  $[0, +\infty]$  such that, for any map  $t \rightarrow s_t$  from  $[0, +\infty[$  to  $\mathbb{S}$  such that  $\lim_{t \rightarrow +\infty} \frac{s_t}{t} = j$ , the limit

$$\lim_{t \rightarrow +\infty} -\frac{1}{t} \ln P(X_t = s_t) \quad (\text{E.18})$$

exists and has value  $g_X(j)$ .

A weaker condition comes from asking whether there is a function  $h_X(j)$  from  $\mathbb{R}$  to  $[0, +\infty]$  such that the limit

$$\lim_{\substack{t \rightarrow +\infty \\ t j \in \mathbb{S}}} -\frac{1}{t} \ln P(X_t = t j) \quad (\text{E.19})$$

exists and has value  $h_X(j)$ .

Of course same care is needed when choosing  $\mathbb{S}$ , which would better be in some way minimal. For instance, if you take  $\mathbb{S} = \mathbb{Z}$  while  $X_t$  takes only even values,  $h_X$  or  $g_X$  have no chance to exist. However, when  $\mathbb{S}$  is chosen appropriately and  $P(X_t = s)$  behaves nicely for large  $t$  and  $s$ , it can be expected on physical grounds that all three functions  $R_X$ ,  $g_X$  and  $h_X$  exist and are the same. Alas, we are not aware of practical and general enough mathematical criteria that guarantee this fact. Of the three,  $h_X$  is probably the easiest to tackle with bare hands. However, to deal with  $R_X$  one can often rely on powerful theorems of the general theory of large deviations. But whether they exist (and coincide) or not, the three functions are meaningful characteristics of the large deviations of  $X_t$ . With some imagination, we could probably invent others.

### E.3 Inequalities at finite time and fluctuation relation between large deviation functions

One important feature of large deviation functions in out-of-equilibrium statistical mechanics is that they satisfy, under appropriate circumstances, symmetry relations. Usually, these relations are consequences of a, possibly generalized, time reversal symmetry. Typically, for some  $0 < m < M < +\infty$  and for some real  $\gamma$  one has

$$mP(X_t = -s)e^{\gamma s} \leq P(X_t = s) \leq MP(X_t = -s)e^{\gamma s} \quad (\text{E.20})$$

for every  $s$ . To be fair, this way of writing things applies, strictly speaking, only when  $X_t$  is atomic, i.e. takes values in a discrete set. A more correct statement is that the laws of  $X_t$  and  $-X_t$  are absolutely continuous and the Radon-Nikodym derivative<sup>5</sup>

$$me^{\gamma s} \leq \frac{dP_{X_t}(s)}{dP_{-X_t}(s)} \leq Me^{\gamma s}, \quad (\text{E.21})$$

of which the previous equation is a special case.

The point we want to make here is that, with any of the three definitions of large deviation functions given above (when they do exist), the above equation implies a symmetry relation. For  $g_X$  or  $h_X$  we can use the simplified equation evaluated when  $s \in \mathbb{S}$ . It implies

$$-\frac{1}{t} \ln M - \gamma \frac{s}{t} - \frac{1}{t} \ln P(X_t = -s) \leq -\frac{1}{t} \ln P(X_t = s) \leq -\frac{1}{t} \ln m - \gamma \frac{s}{t} - \frac{1}{t} \ln P(X_t = -s). \quad (\text{E.22})$$

Assuming the existence of  $h_X$  one can take the limit  $\lim_{\substack{t \rightarrow +\infty \\ tj \in \mathbb{S}}} to get  $h_X(-j) - \gamma j \leq h_X(j) \leq h_X(-j) - \gamma j$ , i.e.$

$$h_X(j) = h_X(-j) - \gamma j, \quad (\text{E.23})$$

the announced symmetry relation. If  $g_X$  exists, then so does  $h_X$  and they are equal, so the symmetry property for  $g_X$  (when it exists) is a consequence of the above bounds. In the case when  $R_X$  exists, the proof of its symmetry based on the inequalities for the Radon-Nikodym derivative is only slightly more elaborate : if  $B$  is a Borel subset of  $\mathbb{R}$  with  $tj_- \leq \inf B$  and  $\sup B \leq tj_+$  then, assuming  $\gamma \geq 0$  for definiteness,

$$P(-X_t \in B)e^{\gamma tj_-} \leq \int_B dP_{-X_t}(s)e^{\gamma s} \leq P(-X_t \in B)e^{\gamma tj_+} \quad (\text{E.24})$$

so, as a consequence of (E.21),

$$-\frac{1}{t} \ln M - \gamma j_+ - \frac{1}{t} \ln P(X_t \in -B) \leq -\frac{1}{t} \ln P(X_t \in B) \leq -\frac{1}{t} \ln m - \gamma j_- - \frac{1}{t} \ln P(X_t \in -B), \quad (\text{E.25})$$

and taking  $B = ]tj_-, tj_+[$  and the  $\limsup_{t \rightarrow +\infty}$  in the second inequality we get

$$\limsup_{t \rightarrow +\infty} -\frac{1}{t} \ln P\left(\frac{X_t}{t} \in ]j_-, j_+[ \right) \leq -\gamma j_- + \limsup_{t \rightarrow +\infty} -\frac{1}{t} \ln P\left(\frac{X_t}{t} \in ]-j_+, -j_-[ \right). \quad (\text{E.26})$$

Similarly, taking  $B = [tj_-, tj_+]$  and the  $\liminf_{t \rightarrow +\infty}$  of the first inequality we get

$$-\gamma j_+ + \liminf_{t \rightarrow +\infty} -\frac{1}{t} \ln P\left(\frac{X_t}{t} \in [-j_+, -j_-] \right) \leq \liminf_{t \rightarrow +\infty} -\frac{1}{t} \ln P\left(\frac{X_t}{t} \in [j_-, j_+] \right). \quad (\text{E.27})$$

Taking  $j_{\pm} = j \pm \varepsilon$ ,  $j$  fixed,  $\varepsilon \searrow 0$  we find from the characterization (E.6) of  $R_X$

$$R_X(j) \leq -\gamma j + R_X(-j) \text{ and } -\gamma j + R_X(-j) \leq R_X(j) \quad (\text{E.28})$$

---

<sup>5</sup>One says that the Radon-Nikodym derivative of a probability measure  $\mu$  with respect to a probability measure  $\nu$  exists if there is an  $\nu$ -integrable function  $f$  such that for every Borel set  $B$  one has  $\int_B d\mu = \int_B f d\nu$ . Then  $f$  is called the Radon-Nikodym derivative of  $\mu$  with respect to  $\nu$  and one writes  $f = d\mu/d\nu$ .

so

$$R_X(j) = R_X(-j) - \gamma j, \quad (\text{E.29})$$

the announced symmetry. So with respect to fluctuation relations, our three definitions of large deviations are on the same footing.

We notice that the above relation is usually written in the more symmetric form  $R_X(j) - R_X(-j) = -\gamma j$ , which is however ambiguous when  $R_X(j)$  is infinite.

#### E.4 Cumulated quantities with sub-extensive difference

We would like to stress that some natural and desirable properties of large deviation functions are not automatic with some of our definitions.

For instance, suppose that  $X_t$  and  $Y_t$  are random processes with a sub-extensive difference, i.e. there is a non-random function  $C_t$  such that  $\lim_{t \rightarrow +\infty} C_t = 0$  and  $|X_t - Y_t| \leq tC_t$ . Then it should be expected that the following alternative holds :

$$\begin{cases} - X_t \text{ and } Y_t \text{ have the same large deviation function} \\ - \text{neither } X_t \text{ nor } Y_t \text{ has a large deviation function.} \end{cases} \quad (\text{E.30})$$

We shall show that this holds true within the general theory of large deviations. With our ad-hoc definitions, we can see that even when  $X_t$  and  $Y_t$  have the required properties, the discrete sets on which  $X_t$  and  $Y_t$  take their values can be very different. And even if the relation is simple, some problems remain at least with the weakest definition of large deviation functions.

To make a proof within the general theory of large deviations, we note that, if  $j_- < j_+$ , any real function  $R$  has the property that

$$\lim_{\varepsilon \searrow 0} \inf_{j \in [j_- - \varepsilon, j_+ - \varepsilon[} R(j) = \inf_{j \in [j_-, j_+[} R(j). \quad (\text{E.31})$$

Moreover, if  $R$  is lower semi-continuous then

$$\lim_{\varepsilon \searrow 0} \inf_{j \in [j_- - \varepsilon, j_+ + \varepsilon]} R(j) = \inf_{j \in [j_-, j_+]} R(j). \quad (\text{E.32})$$

Now, let  $\varepsilon > 0$ . For  $s$  large enough,  $C_s < \varepsilon$  by hypothesis, so that

$$P\left(\frac{Y_s}{s} \in [j_-, j_+]\right) \leq P\left(\frac{X_s}{s} \in [j_- - \varepsilon, j_+ + \varepsilon]\right) \text{ for } s \geq t \quad (\text{E.33})$$

whenever  $t$  is such that  $C_s \leq \varepsilon$  for  $s \geq t$ . Then

$$\inf_{s \geq t} -\frac{1}{s} \ln P\left(\frac{X_s}{s} \in [j_- - \varepsilon, j_+ + \varepsilon]\right) \leq \inf_{s \geq t} -\frac{1}{s} \ln P\left(\frac{Y_s}{s} \in [j_-, j_+]\right), \quad (\text{E.34})$$

so taking  $t \rightarrow +\infty$  one gets

$$\liminf_{t \rightarrow +\infty} -\frac{1}{t} \ln P\left(\frac{X_t}{t} \in [j_- - \varepsilon, j_+ + \varepsilon]\right) \leq \liminf_{t \rightarrow +\infty} -\frac{1}{t} \ln P\left(\frac{Y_t}{t} \in [j_-, j_+]\right). \quad (\text{E.35})$$

Assuming  $X_t$  has a large deviation function, the left-hand side is  $\geq \inf_{j \in [j_- - \varepsilon, j_+ + \varepsilon]} R_X(j)$  by (E.3), so we have proved that for any  $\varepsilon > 0$  one has

$$\inf_{j \in [j_- - \varepsilon, j_+ + \varepsilon]} R_X(j) \leq \liminf_{t \rightarrow +\infty} -\frac{1}{t} \ln P\left(\frac{Y_t}{t} \in [j_-, j_+]\right). \quad (\text{E.36})$$

In the same vein, assuming  $j_- < j_+$ , one has, for  $\varepsilon$  small enough ( $\varepsilon < \frac{j_+ - j_-}{2}$ )

$$P\left(\frac{X_s}{s} \in ]j_- + \varepsilon, j_+ - \varepsilon]\right) \leq P\left(\frac{Y_s}{s} \in ]j_-, j_+[ \right) \text{ for } s \geq t \quad (\text{E.37})$$

whenever  $t$  is such that  $C_s \leq \varepsilon$  for  $s \geq t$ . As above, we infer that

$$\limsup_{t \rightarrow +\infty} -\frac{1}{t} \ln P \left( \frac{Y_t}{t} \in ]j_-, j_+[ \right) \leq \limsup_{t \rightarrow +\infty} -\frac{1}{t} \ln P \left( \frac{X_t}{t} \in ]j_- + \varepsilon, j_+ - \varepsilon[ \right). \quad (\text{E.38})$$

Assuming  $X_t$  has a large deviation function, the right-hand side is  $\leq \inf_{j \in ]j_- + \varepsilon, j_+ - \varepsilon[} R_X(j)$  by (E.4), so for  $\varepsilon$  small enough one has

$$\limsup_{t \rightarrow +\infty} -\frac{1}{t} \ln P \left( \frac{Y_t}{t} \in ]j_-, j_+[ \right) \leq \inf_{j \in ]j_- + \varepsilon, j_+ - \varepsilon[} R_X(j) \quad (\text{E.39})$$

Taking the limit  $\varepsilon \searrow 0$  in (E.36) and (E.39) (using the lower semi-continuity of  $R_X$  in (E.36)) we obtain precisely that  $Y_t$  satisfies an LDP (E.3)–(E.4) with large deviation function  $R_X$ . This proves that the alternative (E.30) holds.

We conclude this discussion with a (counter)-example which has the advantage of being extremely simple, but the drawback that it is artificial (in particular, it is not continuous in probability, it even has non-random discontinuity times). The example is  $X_t \equiv [t]$  the integer part of  $t$ , i.e.  $[t]$  is an integer and  $[t] \leq t < [t] + 1$ .

We urge the reader to check that

-  $X_t$  has a large deviation function  $R_X$  and a large deviation function  $h_X$  which coincide, namely

$$R_X(j) = h_X(j) = \begin{cases} 0 & \text{if } j = 1 \\ +\infty & \text{if } j \neq 1 \end{cases}. \quad (\text{E.40})$$

- If  $Y_t \equiv X_t + 1$  then  $Y_t$  has a large deviation function  $R_Y (= R_X$  by the previous result) but  $h_Y$ , while existing, differs from  $h_X$ , namely

$$h_Y(j) = +\infty \text{ whatever } j. \quad (\text{E.41})$$

-  $X_t$  (resp.  $Y_t$ ) has no large deviation function  $g_X$  (resp.  $g_Y$ ).

Here are proofs for the two first properties. The third is easy and left to the reader.

The case of  $h_X$  : we take  $\mathbb{S} = \mathbb{N}$

$$\lim_{\substack{t \rightarrow +\infty \\ tj \in \mathbb{N}}} -\frac{1}{t} \ln P(X_t = tj) = \lim_{\substack{t \rightarrow +\infty \\ tj \in \mathbb{N}}} -\frac{1}{t} \ln \mathbb{1}_{[t]=tj}. \quad (\text{E.42})$$

If  $j > 1$   $[t] \leq t < tj$  so  $\mathbb{1}_{[t]=tj} = 0$ . If  $j < 1$   $[t] > t - 1 > tj$  whenever  $t > \frac{1}{1-j}$ , so  $\mathbb{1}_{[t]=tj} = 0$  for  $t$  large enough. If  $j = 1$   $\mathbb{1}_{[t]=t} = 1$  for  $t \in \mathbb{N}$ . This gives the announced formula (E.40) for  $h_X$ .

The case of  $R_X$  :  $P \left( \frac{X_s}{s} \in [j_-, j_+] \right) = \mathbb{1}_{[s] \in [sj_-, sj_+]}$ .

- If  $1 \in [j_-, j_+]$  and  $s \in \mathbb{N}$  then  $P \left( \frac{X_s}{s} \in [j_-, j_+] \right) = 1$ , so for any  $t \inf_{s \geq t} -\frac{1}{s} \ln P \left( \frac{X_s}{s} \in [j_-, j_+] \right) = 0$  and  $\liminf_{t \rightarrow +\infty} -\frac{1}{t} \ln P \left( \frac{X_t}{t} \in [j_-, j_+] \right) = 0$ .

- If  $j_- > 1$  then  $[s] < sj_-$  whenever  $s > 0$  so for any  $t > 0 \inf_{s > t} -\frac{1}{s} \ln P \left( \frac{X_s}{s} \in [j_-, j_+] \right) = +\infty$  and  $\liminf_{t \rightarrow +\infty} -\frac{1}{t} \ln P \left( \frac{X_t}{t} \in [j_-, j_+] \right) = +\infty$ .

- If  $j_+ < 1$  then  $[s] > s - 1 \geq sj_+$  whenever  $s \geq \frac{1}{1-j_+}$ , so  $\inf_{s > t} -\frac{1}{s} \ln P \left( \frac{X_s}{s} \in [j_-, j_+] \right) = +\infty$  whenever  $t \geq \frac{1}{1-j_+}$  and  $\liminf_{t \rightarrow +\infty} -\frac{1}{t} \ln P \left( \frac{X_t}{t} \in [j_-, j_+] \right) = +\infty$ .

To summarize

$$\liminf_{t \rightarrow +\infty} -\frac{1}{t} \ln P \left( \frac{X_t}{t} \in [j_-, j_+] \right) = \begin{cases} 0 & \text{if } 1 \in [j_-, j_+] \\ +\infty & \text{else} \end{cases}. \quad (\text{E.43})$$

Analogously  $P \left( \frac{X_s}{s} \in ]j_-, j_+[ \right) = \mathbb{1}_{[s] \in ]sj_-, sj_+[}$ . As  $P \left( \frac{X_s}{s} \in ]j_-, j_+[ \right) \leq P \left( \frac{X_s}{s} \in [j_-, j_+] \right)$  we have

$$\liminf_{t \rightarrow +\infty} -\frac{1}{t} \ln P \left( \frac{X_t}{t} \in ]j_-, j_+[ \right) \leq \limsup_{t \rightarrow +\infty} -\frac{1}{t} \ln P \left( \frac{X_t}{t} \in ]j_-, j_+[ \right). \quad (\text{E.44})$$

As the left-hand side is  $+\infty$  if  $1 \notin [j_-, j_+]$ , so the right-hand side must be  $+\infty$  if  $1 \notin [j_-, j_+]$  and in particular if  $1 \notin ]j_-, j_+[$ . On the other hand, if  $1 \in ]j_-, j_+[$  then  $\lfloor s \rfloor \in ]sj_-, sj_+[$  as soon as  $s \geq \frac{1}{1-j_-}$  so  $\limsup_{t \rightarrow +\infty} -\frac{1}{t} \ln P\left(\frac{X_t}{t} \in ]j_-, j_+[ \right) = 0$ . To summarize

$$\limsup_{t \rightarrow +\infty} -\frac{1}{t} \ln P\left(\frac{X_t}{t} \in ]j_-, j_+[ \right) = \begin{cases} 0 & \text{if } 1 \in ]j_-, j_+[ \\ +\infty & \text{else} \end{cases}. \quad (\text{E.45})$$

From (E.45), (E.43) and the definition of an LDP (E.3)-(E.4) we infer that  $X_t$  has a large deviation function

$$R_X(j) = \begin{cases} 0 & \text{if } j = 1 \\ +\infty & \text{else} \end{cases} \quad (\text{E.46})$$

and  $R_X = h_X$ . The fact that  $X_t$  has no large deviation function  $g_X$  is straightforward.

The case of  $h_Y$  : we take  $\mathbb{S} = \mathbb{N}^*$

$$\lim_{\substack{t \rightarrow +\infty \\ tj \in \mathbb{N}^*}} -\frac{1}{t} \ln P(Y_t = tj) = \lim_{\substack{t \rightarrow +\infty \\ tj \in \mathbb{N}^*}} -\frac{1}{t} \ln \mathbb{1}_{\lfloor t \rfloor + 1 = tj}. \quad (\text{E.47})$$

If  $j > 1$   $\lfloor t \rfloor \leq t+1 < tj$  whenever  $t > \frac{1}{j-1}$  and then  $\mathbb{1}_{\lfloor t \rfloor + 1 = tj} = 0$ . If  $j \leq 1$  then  $\lfloor t \rfloor + 1 > t \geq tj$  for every  $t$  and then  $\mathbb{1}_{\lfloor t \rfloor + 1 = tj} = 0$ . So  $\lim_{\substack{t \rightarrow +\infty \\ tj \in \mathbb{N}^*}} -\frac{1}{t} \ln P(Y_t = tj) = +\infty$  for every  $j$  and  $h_Y(j) = +\infty$  for every  $j$  as announced in (E.41).

## References

- [1] J. L. Lebowitz and H. Spohn. A Gallavotti-Cohen type symmetry in the large deviation functional for stochastic dynamics. *J. Stat. Phys.*, 95:333, 1999.
- [2] L. Onsager. Reciprocal relations in irreversible processes. I. *Phys. Rev.*, 37:405, 1931.
- [3] L. Onsager. Reciprocal relations in irreversible processes. II. *Phys. Rev.*, 38:2265, 1931.
- [4] H.B. Callen. *Thermodynamics*. John Wiley and Sons, 1960.
- [5] S. R. De Groot. *Thermodynamics of Irreversible Processes*. Interscience Publishers, Inc., New York, 1952.
- [6] Y. Oono and M. Paniconi. Steady state thermodynamics. *Prog. Theor. Phys. Suppl.*, 130:29, 1998.
- [7] S. Sasa and H. Tasaki. Steady state thermodynamics. *J. Stat. Phys.*, 125:125, 2006.
- [8] U. Seifert. Stochastic dynamics : principles and perspectives. *Eur. Phys. J. B*, 64:423, 2008.
- [9] U. Seifert. Stochastic thermodynamics, fluctuation theorems and molecular machines. *Rep. Prog. Phys.*, 75:126001, 2012.
- [10] C. Bustamante, J. Liphardt, and F. Ritort. The Nonequilibrium Thermodynamics of Small Systems. *Phys. Today*, 58(7):43, 2005.
- [11] F. Ritort. Single-molecule experiments in biological physics : methods and applications. *J. Phys. Condens. Matter*, 18:R531, 2006.
- [12] K. Mallick. Some recent developments in non-equilibrium statistical physics. *Pramana - J. Phys.*, 73:417, 2009.
- [13] A. Einstein. *Investigations on the Theory of the Brownian Movement*. New York : Dover, 1956.

- [14] U.M.B. Marconi, A. Puglisi, L. Rondoni, and A. Vulpiani. Fluctuation-dissipation : Response theory in statistical physics. *Phys. Rep.*, 461:111, 2008.
- [15] M. Baiesi, C. Maes, and B. Wynants. Fluctuations and response of nonequilibrium states. *Phys. Rev. Lett.*, 103:010602, 2009.
- [16] M. Baiesi, C. Maes, and B. Wynants. Nonequilibrium linear response for Markov dynamics, I. Jump processes and overdamped diffusions. *J. Stat. Phys.*, 137:1094, 2009.
- [17] M. Baiesi, E. Boksenbojm, C. Maes, and B. Wynants. Nonequilibrium linear response for Markov dynamics, II : Inertial dynamics. *J. Stat. Phys.*, 139:492, 2010.
- [18] G. Verley, R. Ch  trite, and D. Lacoste. Modified fluctuation-dissipation theorem for general non-stationary states and application to the Glauber-Ising chain. *J. Stat. Mech.*, page P10025, 2011.
- [19] A. Einstein. Theorie der Opaleszenz von homogenen Fl  ssigkeiten und Fl  ssigkeitsgemischen in der N  he des kritischen Zustandes. *Ann. d. Phys.*, 33:1275, 1910.
- [20] E.T. Jaynes. Information theory and statistical mechanics. *Phys. Rev.*, 106:620, 1957.
- [21] E.T. Jaynes. Information theory and statistical mechanics. II. *Phys. Rev.*, 108:171, 1957.
- [22] R. S. Ellis. *Entropy, large deviations, and statistical mechanics*. Springer, 1985.
- [23] Y. Oono. Large deviation and statistical physics. *Prog. Theor. Phys. Suppl.*, 99:165, 1989.
- [24] A. A. Filyokov and V. Ya Karpov. Method of the most probable path of evolution in the theory of stationary irreversible processes. *Journal of engineering physics and thermophysics*, 13:416, 1967.
- [25] C. Monthus. Non-equilibrium steady states: maximization of the Shannon entropy associated with the distribution of dynamical trajectories in the presence of constraints. *J. Stat. Mech.*, page P03008, 2011.
- [26] D.J. Evans, E.G.D. Cohen, and G.P. Morriss. Probability of second law violations in shearing steady states. *Phys. Rev. Lett.*, 71:2401, 1993.
- [27] D.J. Evans and D.J. Searles. Equilibrium microstates which generate second law violating steady states. *Phys. Rev. E*, 50:1645, 1994.
- [28] G. Gallavotti and E.G.D. Cohen. Dynamical ensembles in nonequilibrium statistical mechanics. *Phys. Rev. Lett.*, 74:2694, 1995.
- [29] G. Gallavotti and E.G.D. Cohen. Dynamical ensembles in stationary states. *J. Stat. Phys.*, page 931, 1995.
- [30] C. Jarzynski. Nonequilibrium Equality for Free Energy Differences. *Phys. Rev. Lett.*, 78:2690, 1997.
- [31] M. Esposito, U. Harbola, and S. Mukamel. Nonequilibrium fluctuations, fluctuation theorems, and counting statistics in quantum systems. *Rev. Mod. Phys.*, 81:1665, 2009.
- [32] M. Campisi, H  nggi P., and P. Talkner. Colloquium : Quantum fluctuation relations : Foundations and applications. *Rev. Mod. Phys.*, 83:771, 2011.
- [33] C. Jarzynski. Equilibrium free-energy differences from nonequilibrium measurements : a master-equation approach. *Phys. Rev. E*, 56:5018, 1997.
- [34] J. Kurchan. Fluctuation theorem for stochastic dynamics. *J. Phys. A: Math. Gen.*, 31:3719, 1998.



- [35] G. E. Crooks. Nonequilibrium measurements of free energy differences for microscopically reversible markovian systems. *J. Stat. Phys.*, 90:1481, 1998.
- [36] G. E. Crooks. Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences. *Phys. Rev. E*, 60:2721, 1999.
- [37] C. Maes. The fluctuation theorem as a Gibbs property. *J. Stat. Phys.*, 95:367, 1999.
- [38] U. Seifert. Entropy production along a stochastic trajectory and an integral fluctuation theorem. *Phys. Rev. Lett.*, 95:040602, 2005.
- [39] M. Esposito and C. Van den Broeck. Three faces of the second law. I. Master equation formalism. *Phys. Rev. E*, 82:011143, 2010.
- [40] B. Derrida. Non-equilibrium steady states: fluctuations and large deviations of the density and of the current. *J. Stat. Mech.*, page P07023, 2007.
- [41] T. Bodineau and B. Derrida. Cumulants and large deviations of the current through non-equilibrium steady states. *C. R. Physique*, 8:540, 2007.
- [42] C. Jarzynski and Wójcik. Classical and quantum fluctuation theorems for heat exchange. *Phys. Rev. Lett.*, 92:230602, 2004.
- [43] D. Andrieux and P. Gaspard. Fluctuation Theorem for Currents and Schnakenberg Network Theory. *J. Stat. Phys.*, 127:107, 2007.
- [44] D. Andrieux and P. Gaspard. Network and thermodynamic conditions for a single macroscopic current fluctuation theorem. *C. R. Physique*, 8:579, 2007.
- [45] F. Cornu and M. Bauer. Thermal Contact II. A Solvable Toy Model. *submitted to J. Stat. Phys.*, 2013.
- [46] G. E. Crooks. Path-ensemble averages in systems driven far from equilibrium. *Phys. Rev. E*, 61:2361, 2000.
- [47] D. R. Cox and H. D. Miller. *The Theory of Stochastic Processes*. Chapman and Hall, London, 1965.
- [48] N.G. van Kampen. *Stochastic processes in physics and chemistry*. North Holland, 1992.
- [49] J. T. Lewis and R. Russell. An introduction to large deviations for teletraffic engineers. *Unpublished, to be found on the web*, 1997.
- [50] S. R. S. Varadhan. Large deviations. *The Annals of Probability*, 36:307, 2008.
- [51] D. Andrieux and P. Gaspard. A fluctuation theorem for currents and non-linear response coefficients. *J. Stat. Mech.*, page P02006, 2007.
- [52] J. Schnakenberg. Network theory of microscopic and macroscopic behavior of master equation systems. *Rev. Mod. Phys.*, 48:571, 1976.
- [53] F. Schlögl. On Stability of Steady States. *Z. Phys.*, 243:303, 1971.
- [54] S. Kullback and R. A. Leibler. On information and sufficiency. *Ann. Math. Stat.*, 22:79, 1951.
- [55] N. Pottier. *Physique statistique hors équilibre, Processus irréversibles linéaires*. EDP Sciences, CNRS ÉDITIONS, 2007.
- [56] E.T. Jaynes. The minimum entropy production principle. *Ann. Rev. Phys. Chem.*, 31:579, 1980.

- [57] V. Lecomte, C. Appert-Rolland, and F. van Wijland. Chaotic properties of systems with Markov dynamics. *Phys. Rev. Lett.*, page 010601, 2005.
- [58] I. Prigogine. *Introduction to thermodynamics of irreversible processes*. Wiley, New York, 1968.
- [59] C. Jarzynski. Hamiltonian derivation of a detailed fluctuation theorem. *J. Stat. Phys.*, 98:77, 2000.
- [60] G.N. Bochkov and Yu.E. Kuzovlev. Nonlinear fluctuation-dissipation relations and stochastic models in nonequilibrium thermodynamics I. Generalized fluctuation-dissipation theorem. *Physica A*, 106:443, 1981.
- [61] G.N. Bochkov and Yu.E. Kuzovlev. Nonlinear fluctuation-dissipation relations and stochastic models in nonequilibrium thermodynamics II. Kinetic potential and variational principles for nonlinear irreversible processes. *Physica A*, 106:480, 1981.
- [62] M. Esposito and C. Van den Broeck. Three detailed fluctuation theorems. *Phys. Rev. Lett.*, 104:090601, 2010.
- [63] H. Touchette. The large deviation approach to statistical mechanics. *Phys. Rep.*, 478:1, 2009.
- [64] A. Dembo and O. Zeitouni. *Large Deviations Techniques and Applications*. Springer, New York, 2nd edition, 1998.
- [65] E. Helfand. Transport coefficients from dissipation in a canonical ensemble. *Phys. Rev.*, 119:1, 1960.
- [66] M. S. Green. Markoff random processes and the statistical mechanics of time-dependent phenomena. *J. Chem. Phys.*, 20:1281, 1952.
- [67] M.S. Green. Markoff random processes and the statistical mechanics of time-dependent phenomena. II irreversible processes in fluids. *J. Chem. Phys.*, 22:398, 1954.
- [68] R. Kubo. Statistical-mechanical theory of irreversible processes. I. General theory and simple applications to magnetic and conduction problems. *J. Phys. Soc. Japan*, 12:570, 1957.
- [69] R. Kubo, M. Yokota, and S. Nakajima. Statistical-mechanical theory of irreversible processes. II. Response to thermal disturbance. *J. Phys. Soc. Japan*, 12:1203, 1957.
- [70] T. Yamamoto. Quantum statistical mechanical theory of the rate of exchange chemical reactions in the gas phase. *J. Chem. Phys.*, 33:281, 1960.
- [71] R. Zwanzig. Time-correlation functions and transport coefficients in statistical mechanics. *Ann. Rev. Phys. Chem.*, 16:67, 1965.
- [72] A. Faggionato and D. Di Pietro. Gallavotti-Cohen-type symmetry related to cycle decompositions for Markov chains and biochemical applications. *J. Stat. Phys.*, 143:11, 2011.
- [73] D. Andrieux and P. Gaspard. Fluctuation theorem and Onsager reciprocity relations. *J. Chem. Phys.*, 121:6167, 2004.
- [74] P. E. Wigner. Derivations of Onsager’s reciprocal relations. *J. Chem. Phys.*, 22:1912, 1954.